

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: May 26, 2003, 13:09:27 ; Search time 21 Seconds

(Without alignments)  
667.571 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1789

Sequence: 1 MGRKSKVRLALRSPKRG.....SRIVPQRLRKQPVRRKQ 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1789	100.0	338	1	DEFB_HUMAN
2	1416	79.2	344	1	DEFB_MOUSE
3	1392	77.6	349	1	DEFB_RAT
4	108.5	6.1	1920	1	RNT_MOUSE
5	98	5.3	448	1	VIME_MOUSE
6	98	5.3	448	1	VIME_CRIGR
7	98	5.3	464	1	VIME_MESAU
8	98	5.5	465	1	VIME_RAT
9	97	5.4	465	1	VIME_BOVIN
10	96	5.4	465	1	VIME_HUMAN
11	94.5	5.3	1108	1	DBS_HUMAN
12	92.5	5.2	1054	1	FS27_HUMAN
13	92.5	5.1	841	1	POL_STYAK
14	91.5	5.1	841	1	BAC2_HUMAN
15	91	5.1	459	1	VIME_CHICK
16	90.5	5.1	1022	1	POL_STYSP
17	90.5	5.1	1149	1	DBS_MOUSE
18	90	5.0	1211	1	ATC6_SCHPO
19	89.5	5.0	763	1	DESM_XENLA
20	88.5	4.9	458	1	DESM_RABIT
21	88.5	4.9	1019	1	DESM_XENLA
22	88.5	4.9	1035	1	POL_HV28B
23	88.5	4.9	1073	1	POL_HV28B
24	88.5	4.9	1738	1	CO4_MOUSE
25	87.5	4.9	1042	1	POL_HV28B
26	87.5	4.9	1249	1	RO2_ASPP7
27	87	4.9	1066	1	SPS5_SCHPO
28	87	4.9	1066	1	SPS5_SCHPO
29	86.5	4.8	3321	1	PCN2_HUMAN
30	86.5	4.8	219	1	CIDB_HUMAN
31	86	4.8	219	1	CIDB_HUMAN
32	86	4.8	458	1	VIMI_XENLA
33	86	4.8	463	1	VIMI_XENLA

## ALIGNMENTS

34	86	4.8	1027	1	KINN_MOUSE
35	85.5	4.8	331	1	DEFA_MOUSE
36	85.5	4.8	736	1	XMS1_DROME
37	85.5	4.8	1744	1	CO4_HUMAN
38	84.5	4.7	373	1	CONS_ARATH
39	84.5	4.7	1056	1	POL_STYML
40	84.5	4.7	3859	1	ROPA_LELV
41	84	4.7	217	1	CIDB_MOUSE
42	84	4.7	219	1	CIDB_MOUSE
43	84	4.7	526	1	GSPD_CEREA
44	83.5	4.7	239	1	FS27_MOUSE
45	83.5	4.7	308	1	YDM7_SCHPO

  

RESULT 1	DEFB_HUMAN	STANDARD;	PRT; 338 AA.
ID	DEFB_HUMAN	O60521; Q9BYT4; Q9BYT6;	
AC	O60521; Q9BYT4; Q9BYT6;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DE	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	DNA fragmentation factor 40 kDa subunit (EC 3.-.-.) (DEF-40)		
DE	(Caspase-activated deoxyribonuclease) (Caspase-activated DNase) (CAD)		
DE	(Caspase-activated nuclease) (CPAN).		
GN	DEFB OR DEF2 OR DEF40 OR CAD.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
NC	NCBI_TaxID=9606;		
LN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA).		
RX	MEDLINE=98337937; PubMed=9671700;		
RA	Liu X., Li P., Widlak P., Zou H., Luo X., Garrard W.T., Wang X.;		
RT	"The 40-kDa subunit of DNA fragmentation factor induces DNA		
RT	fragmentation and chromatin condensation during apoptosis.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 93:8461-8466(1996).		
RT	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA).		
RX	MEDLINE=98336121; PubMed=9689044;		
RA	Murata N., Emari M., Sakahira H., Fukuda Y., Inazawa J., Toh H.,		
RT	Nagata S.;		
RT	"Molecular cloning and characterization of human caspase-activated		
RT	DNase.";		
RT	Proc. Natl. Acad. Sci. U.S.A. 95:9123-9128(1998).		
RT	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA).		
RX	TISSUE-Pancreas;		
RA	MEDLINE=98228358; PubMed=9560346;		
RT	Halenebeck R., Macdonald H., Roulston A., Chen T.T., Conroy L.,		
RT	Williams L.T.;		
RT	"CPAN, a human nuclease regulated by the caspase-sensitive inhibitor		
RT	DEF45.";		
RT	Curr. Biol. 8:537-540(1998).		
RT	[4]		
RP	SEQUENCE FROM N.A. (ISOFORMS BETA; GAMMA AND DELTA).		
RC	TISSUE-Fetal brain;		
RA	Nakagawa A., Takahashi M., Takada N., Kawamoto T.;		
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.		
RT	[5]		
RP	VARIANT LXS-196		
RX	MEDLINE=20283753; PubMed=10830907;		
RA	Uudson H., Van Roy N., Strain U., Vandesompele J., Van Gele M.,		
RT	Speelman F., Boulton D.T.;		
RT	Structure and mutation analysis of the gene encoding DNA		
RT	fragmentation factor 40 (caspase-activated nuclease), a candidate		
RT	neuroblastoma tumour suppressor gene.";		
RT	Hum. Genet. 106:406-413(2000).		
CC	1- FUNCTION: Nuclease that induces DNA fragmentation and chromatin		
CC	condensation during apoptosis. Degrades naked DNA and induces		
CC	apoptotic morphology.		



[illegible]

```

OY      181 EAOGEFFLRVAGSCMOGLRSMQNGSGYDPRACAKGSLCTPCEWESCOGPPDMOSCLSRHS 240
Db      184 AAREEYLRLVSNCHKLKSVQYNSSGYDRGAEMASHLCTPBEWESSCGPPDELSCLSKHS 243
OY      241 IMPNSNEPSRLSTPNNDLHTEKKRTTITLYVEALKEBDQREVMDFEGILLFSENLK 300
Db      244 IMPGNRESRLSTPNNDLHTEKKRTTVPLLEAL--ODGNEVMEWFSLTLFAENLK 301
OY      301 LVHIYCHKTKTHLNCDSRYIKPOTRIKKRPVKRR 337
Db      302 LVHIACHKTKTHLQCDSRLTYRPQTGSRRKPVRKO 338

RESULT 4
PCNT_MOUSE STANDARD; PRT: 1920 AA.
ID PCNT_MOUSE
AC PA8725;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Pericentrin.
GN PCNT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
MM Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX MGI|taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=94170365; PubMed=8124707;
RX Dosey S.J., Stein P., Evans L., Calarco P.D., Kirschner M.;
RT "Pericentrin, a highly conserved centrosome protein involved in
   microtubule organization.";
RL Cell 76:639-650(1994).
CC -1- FUNCTION: INTEGRAL COMPONENT OF THE FILAMENTOUS MATRIX OF THE
   CENTROSOME INVOLVED IN THE INITIAL ESTABLISHMENT OF ORGANIZED
   MICROTUBULE ARRAYS IN BOTH MITOSIS AND MEIOSIS.
CC CC TISSUE SPECIFICITY: HIGHEST LEVELS FOUND IN KIDNEY, THYMUS AND
   LIVER. LOW LEVELS IN BRAIN, MUSCLE, LUNG AND HEART.
CC -1- DOMAIN: COMPOSED OF A COILED-COIL CENTRAL REGION FLANKED BY NON-
   HELICAL N- AND C-TERMINALS.
CC -1- SIMILARITY: STRONG. TO HUMAN KENDRIN.
Cc -----
Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL Outstation -
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb.ch/announce/
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc DR EMBL; 005823; AAA17886.1; -.
Cw MGD; MGI:102722; Pcnt.
KW Coiled coil; Microtubules.
FT DOMAIN 110 1600 COILED COIL (POTENTIAL).
FT FT 109 112 POLY-PRO.
FT FT 383 387 POLY-GLN.
SQ SEQUENCE 1920 AA; 218337 MW; CF1DDADECSB73309 CRG64;

Query Match 6.1%; Score 108.5; DB 1; Length 1920;
Best Local Similarity 23.9%; Pred. No 0.48; Indels 43; Gaps 11;
Matches 56; Conservative 45; Mismatches 11;

OY      2 LOKRYSYKIALSLSPRKREIVAGRSQEVLT-RKC---LRQLPEGRSLCTLY--ED 51
Db      142 LOKERETRLTELKEM-----LNGRRROELTALQSQQCELELTRDHAREKEMALSSGGE 197
OY      52 GTLETEDFPVSVDNAELVLLTLGAWOG-----YSDIRRFILSAFHQPVLQAQOULL 107
Db      198 LAELKEKRLSEMKNQ-Q-TIETLKQDWSEIRELCLENLOELSLKHOSEWEGJOSOFQRE 256
OY      108 CDQDAPOHORLL-----ADLIHWANSONIA-----ETRAEDPFWFEGLSEFQSGK 154

```

Db 257 L5EQKVELEKTFQKKHAEVSLKNLEAQHOQA1KKLQEDLQSHQCYLDQJEDKFRFKFKE

QY 155 YLRSCSRIRSYLREYSSPTVYGAAQOEFLRVLSGMOQRARS--MQYNGSY 206

Db 317 AKELLETTLQASY-----EDLKAOSE-IRLLMSQLESMTKTRREELNGSW 361

RESULT 5

VIME\_MOUSE

ID VIME\_MOUSE STANDARD; PRT; 465 AA.

AC P20152; 008704;

DT 01-FEB-1991 (Rel. 17, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Vimentin.

GN VIM.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NCBI\_Taxid=10090;

[1]

SEQUENCE FROM N.A.

MEDLINE=69306653; PubMed=2744479;

Wood L., Thieriault N., Vogel G.;

"Vimentin cDNA clones covering the complete intermediate-filament protein are found in an EHS tumor cDNA library.";

Gene 76:171-175(1989).

[2]

SEQUENCE FROM N.A.

STRAIN=C57BL/6; TISSUE=Spleen;

Podolin P.L., Prystowsky M.B.;

Submitted (OCT-1990) to the EMBL/GenBank/DBJ databases.

[3]

SEQUENCE FROM N.A.

MEDLINE=90220517; PubMed=2325630;

Hennekes H., Kuehn S., Tirub P.;

"Coding sequence and flanking regions of the mouse vimentin gene.";

Mol. Gen. Genet. 221:33-36(1990).

[4]

SEQUENCE FROM N.A.

TISSUE=Smooth muscle;

MEDLINE=90265604; PubMed=2140597;

Capetanaki Y., Kulak I., Rothblum K., Starnes S.;

"Mouse vimentin: structural relationship to fos, jun, CREB and tpr.";

Oncogene 5:645-655(1990).

[5]

SEQUENCE FROM N.A.

Rauscher A.;

Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.

[6]

SEQUENCE OF 1-114 FROM N.A.

STRAIN-BALB/C;

MEDLINE=96125204; PubMed=8543176;

Nakamura N., Shida M., Hirayoshi K., Nagata K.;

"Transcriptional regulation of the vimentin-encoding gene in mouse myeloid leukemia M1 cells.";

Gene 166:281-286(1995).

[7]

SEQUENCE OF 71-90.

TISSUE=Fibroblast;

MEDLINE=95009907; PubMed=7523108;

Merrick B.A., Patterson R.M., Wichter L.L., He C., Selkirk J.K.;

"Separation and sequencing of a familiar and novel murine proteins using preparative two-dimensional gel electrophoresis.";

Electrophoresis 15:735-745(1994).

[8]

PHOSPHORYLATION OF SER-38 AND SER-82.

MEDLINE=91222208; PubMed=1850597;

Ando S., Tokui T., Yamauchi T., Sugura H., Tanabe K., Inagaki M.;

"Evidence that Ser-82 is a unique phosphorylation site on vimentin for Ca2(+)-calmodulin-dependent protein kinase II.";

Biochem. Biophys. Res. Commun. 175:955-962(1991).

-1- FUNCTION: Vimentins are class-III intermediate filaments found in

Query Match	Best Local Similarity	Score	DB 1	Length	465
Matches 63; Conservative	47; Mismatches 113; Indels 90; Gaps 14				
27	QEVLRKGRQQLPQERGRSLCLYEDGTALTEDYFSPVDNAELVLTITGQAGGYVSDIR	86			
179	EDIMR---LRKRLQEE---MIGREASTYLOSFGQVDNMSLADLDERKYESIQEEL-	230			
QY	87 RFLSAFHEPQVGLIOA-----QQLCDQAPQORRLIADLIHNSONTAETRAEDPPW	141			
Db	231 AFLKRLHDEETQELQAOIQEOHVQIDVVSXPDLTALRDVFOQY-ESVAANKILOAEVEM	289			
QY	142 F-----EGLESRFQSKGYLR-----SCESRIR-----S	166			
Db	290 YKSFADLSKANNRNDALRQAKQSNETRRVOVSLTCEVDALKTNSLERQKREMEEN	349			
QY	167 YLREVSSTPVGAENAEFLRVLSMCOURLRSMQYNGSYEDRGAKGSRLLCTPGWFSQ	226			
Db	350 FALFANAYQDITIG-RLQDEIQNMKREMARHLREYQ-----DLNV	388			
QY	227 QGPDMQSCLSRHSINPISNRSR-----LFTWNLDTHT-----IEKKRTIIPPL	272			
Db	389 KMAIDIEATYRKLK---EGEESRSTSLPFPFSSINLTETNLSLPLVDYTHSKRL- -L	442			

QY	273	VEAIKEDQGREVD	285
Db	443	IKVETREDQGVIN	455
RESULT 6			
ID	VIMC_CRIGR	STANDARD:	PRT: 448 AA.
AC	P48670		
DT	01-FEB-1996	(Rel. 33, Created)	
DT	01-FEB-1996	(Rel. 33, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Vimentin (Fragment).		
GN	VIM.		
OS	Cricetulus griseus (Chinese hamster).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;		
OS	Cricetulus.		
OS	NCBI_TaxID=10029;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=83297273; PubMed=668458;		
RA	Bioemondal H., Quax W., Quax-Jeukens Y., Dodemont H., Ramaekers F.,		
RA	Dunja I., Benedetti L.;		
RT	"Organization and expression of the vimentin gene.";		
RL	Mol. Biol. Rep. 9:115-118(1983).		
RP	SEQUENCE OF 69-448 FROM N.A.		
RA	Rhodolajov A.L., Koonce M.P.;		
RA	Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.		
CC	-1- FUNCTION: Vimentins are class-III intermediate filaments found in		
CC	various non-epithelial cells, especially mesenchymal cells.		
CC	-1- SUBUNIT: HOMOPOLYMER.		
CC	-1- PTM: ONE OF THE MOST PROMINENT PHOSPHOPROTEINS IN VARIOUS CELLS OF		
CC	MESENCHYMAL ORIGIN. PHOSPHORYLATION IS ENHANCED DURING CELL		
CC	DIVISION, AT WHICH TIME VIMENTIN FILAMENTS ARE SIGNIFICANTLY		
CC	REORGANIZED (BY SIMILARITY).		
CC	-1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See <a href="http://www.isb-sb.ch/announce/">http://www.isb-sb.ch/announce/</a>		
CC	or send an email to <a href="mailto:license@isb-sb.ch">license@isb-sb.ch</a> ).		
CC	-----		
CC	EMBL: M16718; AAA37029.1; -		
CC	EMBL: X87227; CAA60678.1; -		
CC	InterPro: IPR001664; IF: -		
DR	Pham: PF00036; Filament; 1.		
DR	PROSITE: PS00226; IF: 1.		
FM	Intermediate filament; Coiled coil; Phosphorylation.		
FT	NON-TER		
FT	FT DOMAIN	1	77
FT	FT DOMAIN	78	389
FT	FT DOMAIN	78	389
FT	FT DOMAIN	390	448
FT	FT DOMAIN	78	113
FT	FT DOMAIN	114	135
FT	FT DOMAIN	136	227
FT	FT DOMAIN	228	250
FT	FT DOMAIN	251	389
FT	FT DOMAIN	COIL 2.	
SO	SEQUENCE	448 AA; 51848 MW; 251880FE759B3B87 CRC64;	
Query Match		5.5%; Score 98; DB 1; Length 448;	
Best Local Similarity		20.1%; Pval. No. 0.59;	
Matches	63; Conservative	47; Mismatches	113; Indels
			90; Gaps
14; Gaps			
27	OEVIARKCCTAFQPEGSGSLCTLYEGFLEDEDFSVPNALVLTITGQAMQGYSDIR	86	
162	EDIIIR---LREKIQEE---MLQREASTLQSGFRODVNMSLARLDERKVESLQEEI-	213	

```

QY 87LESLAEFEPOVGLIOAA-----OOLTCDEQAPORQRLADLLHNSONIAAEFRADBPW 141
Db 214 AFLKKLHDEEITQELAQAOIOEHOVIDVSKPDLTALADVYNOQ-ESVAKKMLQGAEM 272
OY 142 F-----EGESRFOSKSGYRY-----SCESRIR-----S 166
Db 273 YKSNFADLSAARNNDALRQAKQESNENYRQVQSLTCEVDALNGTNELEQRMREBN 332
OY 167 YLAEVSSPTVGAEOEFLRYLQSCQRLBENQVNSYFDRGAKGSLRCLPBEWFC 226
Db 333 FALFNAYQOTIG-RLODIDQMKKEFAKHLREXO-----DLLNV 371
OY 227 OGCPDMQCSLRHSINPNSRSL-----STWALDHI-----IEKRTIIPPL 272
Db 372 KALDIDIAFYRKL-----EGESRISLPLPNESSLNLTRETNESLPLVDTHSKRTL--L 425
OY 273 YEAKIEDQGREVD 285
Db 426 IKVETRDGOVIN 438

RESULT 7
VIME_MESAU STANDARD; PRT; 464 AA.
AC P02544;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Vimentin.
GN VIM.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognath; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
[1]
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]
[520]
[521]
[522]
[523]
[524]
[525]
[526]
[527]
[528]
[529]
[530]
[531]
[532]
[533]
[534]
[535]
[536]
[537]
[538]
[539]
[540]
[541]
[542]
[543]
[544]
[545]
[546]
[547]
[548]
[549]
[550]
[551]
[552]
[553]
[554]
[555]
[556]
[557]
[558]
[559]
[560]
[561]
[562]
[563]
[564]
[565]
[566]
[567]
[568]
[569]
[570]
[571]
[572]
[573]
[574]
[575]
[576]
[577]
[578]
[579]
[580]
[581]
[582]
[583]
[584]
[585]
[586]
[587]
[588]
[589]
[590]
[591]
[592]
[593]
[594]
[595]
[596]
[597]
[598]
[599]
[600]
[601]
[602]
[603]
[604]
[605]
[606]
[607]
[608]
[609]

```

```

DR EMBL; K00925; AAA37104.1; JOINED.
DR EMBL; K00926; AAA37104.1; JOINED.
DR PIR; A02959; VEHY.
DR InterPro: IPRO01664; IF.
DR Pfam: PF00038; filament.1.
DR PROSITE; PS00226; IF.1.
KW Intermediate filament; Coiled coil; Acetylation; Phosphorylation.
FT INT_MET 0
FT DOMAIN 1 93 HEAD.
FT DOMAIN 94 405 ROD.
FT DOMAIN 406 464 TAIL.
FT DOMAIN 94 129 COIL 1A.
FT DOMAIN 130 151 LINKER 1.
FT DOMAIN 152 243 COIL 1B.
FT DOMAIN 244 266 LINKER 12.
FT DOMAIN 267 405 COIL 2.
FT MOD_RES 1 1 ACETYLATION.
FT MOD_RES 38 38 PHOSPHORYLATION (BY CAMK2) (BY SIMILARITY).
MOD_RES 81 81 PHOSPHORYLATION (BY CAMK2) (BY SIMILARITY).
CONFLICT 42 42 L -> A (IN REF. 2).
CONFLICT 115 115 Y -> D (IN REF. 2).
CONFLICT 182 182 R -> I (IN REF. 2).
FT SEQUENCE 464 AA; 53598 MW; 7AC417008C8D4776 CRC64;

Query Match 5.5%; Score 98; DB 1; Length 464;
Best Local Similarity 20.1%; Pred. No. 0.62;
Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

QY 27 QEVLRKGLRQLEPGRSRLCLYEDGTELTEDYPPSVNDNELVLTGAMQGVSDIR 86
DB 178 EDIMR---LRKRLQEE---MLQREAEASTIQSFROVDNASTLARLDERVESIQEETI 229
QY 87 RFLSAFHEPQVGLIQAA-----QQLLCDEQAPQORRLADLLHNVSONIAETRAEDPPW 141
DB 230 AFLKKHDEEIOEQAOIQEOHQVIDVDSKPDIALRDVROQY-ESVAANKIQEAEW 288
QY 142 F-----EGLESRFQSKGYLRY-----SCESRIR---S 166
DB 289 YKSKFADISEANNNNDLRQAQESNEYRQVSLCEVDALKTGNTESLEROMREEMEN 348
QY 167 YLREVSYPSTVGAQAEFLRVLSGMCQRLRSQYNGSYDRAGKGSRLCTEGWFS 226
DB 349 FALPAAANTQDTIG-RLDDEIQNMKEEAKRLREYO-----DILLNV 387
QY 227 QGPPDMSCLSRHSINPYSNRESRL-----FSTWNLDHI-----IEKKRTIIPTL 272
DB 388 KMAIDIEIATYRKLL---EGESRISLPLPWFSSILNRETNLJESLPLVDTHSKRTL---L 441
273 VEAIKEDGGRVD 285
442 IKVETRDGQVIN 454

```

```

RA Debruyne F.M.J., Schalken J.A.;
RT "Differential expression of vimentin in rat prostatic tumors.";
RL Biochem. Biophys. Res. Commun. 182:1254-1259(1992).
RN [2]
RP SEQUENCE OF 85-159 FROM N.A.
RA Paine M.L.;
RL Submitted (xxx-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Vimentins are class-III intermediate filaments found in
CC various non-epithelial cells, especially mesenchymal cells.
CC -1- SUBUNIT: HOMOPOLYMER.
CC -1- PTM: ONE OF THE MOST PROMINENT PHOSPHOPROTEINS IN VARIOUS CELLS OF
CC MESENCHYMAL ORIGIN. PHOSPHORYLATION IS ENHANCED DURING CELL
CC DIVISION, AT WHICH TIME VIMENTIN FILAMENTS ARE SIGNIFICANTLY
CC REORGANIZED.
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X62952; CAA44722.1; -.
DR EMBL; M84481; AAA42339.1; -.
DR PIR; S22119; S22119.
DR PIR; J01389; J01389.
DR InterPro: IPRO01664; IF.
DR Pfam: PF00038; filament.1.
DR PROSITE; PS00226; IF.1.
KW Intermediate filament; Coiled coil; Phosphorylation.
FT INT_MET 0
FT DOMAIN 1 94 HEAD.
FT DOMAIN 95 406 ROD.
FT DOMAIN 407 465 TAIL.
FT DOMAIN 95 130 COIL 1A.
FT DOMAIN 131 152 LINKER 1.
FT DOMAIN 153 244 COIL 1B.
FT DOMAIN 245 267 LINKER 12.
FT DOMAIN 268 406 COIL 2.
FT MOD_RES 38 38 PHOSPHORYLATION (BY CAMK2) (BY SIMILARITY).
FT MOD_RES 82 82 PHOSPHORYLATION (BY CAMK2) (BY SIMILARITY).
FT SEQUENCE 465 AA; 53601 MW; 6587FE7652CB3682 CRC64;

Query Match 5.5%; Score 98; DB 1; Length 465;
Best Local Similarity 20.1%; Pred. No. 0.62;
Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

QY 27 QEVLRKGLRQLEPGRSRLCLYEDGTELTEDYPPSVNDNELVLTGAMQGVSDIR 86
DB 179 EDIMR---LRKRLQEE---MLQREAEASTIQSFROVDNASTLARLDERVESIQEETI 230
QY 87 RFLSAFHEPQVGLIQAA-----QQLLCDEQAPQORRLADLLHNVSONIAETRAEDPPW 141
DB 231 AFLKKHDEEIOEQAOIQEOHQVIDVDSKPDIALRDVROQY-ESVAANKIQEAEW 289
QY 142 F-----EGLESRFQSKGYLRY-----SCESRIR---S 166
DB 290 YKSKFADISEANNNNDLRQAQESNEYRQVSLCEVDALKTGNTESLEROMREEMEN 349
QY 167 YLREVSYPSTVGAQAEFLRVLSGMCQRLRSQYNGSYDRAGKGSRLCTEGWFS 226
DB 350 FALPAAANTQDTIG-RLDDEIQNMKEEAKRLREYO-----DILLNV 388
QY 227 QGPPDMSCLSRHSINPYSNRESRL-----FSTWNLDHI-----IEKKRTIIPTL 272
DB 389 KMAIDIEIATYRKLL---EGESRISLPLPWFSSILNRETNLJESLPLVDTHSKRTL---L 442
273 VEAIKEDGGRVD 285
442 IKVETRDGQVIN 454

```

Db 443 IKTETRODOGVIN 455

RESULT 9

ID	TIME	HUMAN	STANDARD:	PRT:	465 AA.
AC	AD	P08670;	Q15867; O9NTM3; Q15869; Q15868;		
AD	AD	01-JAN-1988	(rel. 06, Created)		
DT	01-JUN-1993	(rel. 27, Last sequence update)			
DT	15-JUN-2002	(rel. 41, Last annotation update)			
DE		vimentin.			
DE		vimentin.			
OS		Homo sapiens (Human).			
OC		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI		Taxid:9606;			
NCBI		Taxid:9606;			
SEQUENCE FROM N.A.					
MEDLINE-87089701;		PubMed-3467175;			
Ferrari S., Battini R., Kaczmarek L., Rittling S., Calabretta B.,					
de Riel J.R., Philippou V., Wei J.-F., Baserga R.,					
"Coding sequence and growth regulation of the human vimentin gene.";					
Mol. Cell. Biol. 6:3614-3620(1986).					
[2]					
SEQUENCE FROM N.A.					
MEDLINE-91067467;		PubMed-2251132;			
Honore B., Madsen P., Basse B., Andersen A., Walbum E., Cells J.E.,					
Leffers H.,					
"Nucleotide sequence of cDNA covering the complete coding part of the					
human vimentin gene.";					
Nucleic Acids Res. 18:6692-6692(1990).					
[3]					
SEQUENCE FROM N.A.					
Zimbelmann R.;					
Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.					
[4]					
SEQUENCE FROM N.A.					
Bird C.;					
Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.					
[5]					
SEQUENCE FROM N.A.					
Tissue-Cervix;					
Straussberg R.;					
Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.					
[6]					
SEQUENCE OF 1-134 FROM N.A., AND TISSUE SPECIFICITY.					
TISSUE-Breast carcinoma;					
MEDLINE-89303836;		PubMed-2472876;			
Sommers C.L., Walker-Jones D., Heckford S.E., Worland P.,					
Valverius E., Clark R., McCormick F., Stampfer M., Ablarath S.,					
Yelaman E.P.;					
"Vimentin rather than keratin expression in some hormone-independent					
breast cancer cell lines and in oncogene-transformed mammary					
epithelial cells.";					
Cancer Res. 49:4258-4263(1989).					
[7]					
SEQUENCE OF 112-465 FROM N.A., AND TISSUE SPECIFICITY.					
TISSUE-Fibroblast;					
MEDLINE-88226018;		PubMed-3371665;			
Perran J., Lilienbaum A., Vasseur M., Paulin D.,					
"Nucleotide sequence of the human vimentin gene and regulation of its					
transcription in tissues and cultured cells.";					
Gene 62:7-16(1988).					
[8]					
SEQUENCE OF 166-465 FROM N.A.					
TISSUE-Osteosarcoma;					
MEDLINE-90215314;		PubMed-2323579;			
Gupta A.K., Adin J.E., Wayne M.W.Y.;					
"Isolation of a human vimentin cDNA with a long 3'-noncoding region					
from a human osteosarcoma cell line (MG-63).";					
Gene 86:303-304(1990).					
[9]					
SEQUENCE OF 16-24 AND 54-69.					







```

Best Local Similarity 21.7%; Pred. No. 4.8;
Matches 93; Conservative 38; Mismatches 136; Indels 161; Gaps 21.

OY 15 SPRKRGVAGSCOE-----VLKRGCLR-----FQLRPG-----SRCLVEDGT 53
Db 311 ATPTAPAGAACLCERSVAVSPSCRSLFTSTKSVLSLSPSTSGOHPANSPACPDKCI 376
OY 54 -----ELTEDEPSVPD-----NAETVILTLGAANG----- 80
Db 377 TGGDLKTDYPTFGNIGCGPHVCGKQVNSFTGSLPLGPGELMCKQSEELDRSVTFSS 436
OY 81 -----YVDIRFLISAFHEP-QVGLQAQQLCDDEQAPQORLADLL--- 123
Db 437 ACDQVSTSVHSYSGVSLDKDLS---EVPKGLMVGAGSLPSSQAVSHGELMADHLPCR 493
OY 124 ---HNVSQNIATETRADPWFEEGLETSRFSKSYLRYSCESRINSY-----L 168
Db 494 MKPNKSCVPLKVCPRSP---LETRIRIS-----SCSS---YSYADSGSGPSCLPL 542
OY 169 REVSSPYTVGA-----EAOEERFLVGLSGMCORLRSMQVNSGYFDRGAKGSRICPEGM 223
Db 543 CEFSSSPSCGSAFELATEHQEPGLMGDMYNOVRPOIKCEQSY-----GTNSDEGSS 595
OY 224 FSCGPFMDSCISRSRISIPYNSRESRLTSTWNLDHIEKRTIPIPLVEAKEDQGRE 283
Db 596 ES---EADSESC-----PVQDRQGEVKL-PPVQJITDLPNDPQMIMKMK----- 638
OY 284 VDMEEYFGLLEFTESENKLTHIV-----CHKK-----TTHKLCDPSRI 321
Db 639 -----LTSEQLEFIDVRRRSKRNRIAAQRCKRRKLDIQNLBEICRIKLVCKERKL 688
OY 322 YKPPORLK 329
Db 689 LSRNQOLK 696

RESULT 15
VIME.CHICK ID VIME.CHICK STANDARD: PRT: 459 AA.
AC P09654; 091023;
Db 01-MAR-1989 (Rel. 10, Created)
Db 01-MAR-1989 (Rel. 10, Last sequence update)
Db 15-JUN-2002 (Rel. 41, Last annotation update)
Db Vimentin.
VIM.
Gallus gallus (Chicken).
Archaeopteryx; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archaeosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
Gallus.
NCBI_TaxID=9031;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=87250401; Pubmed=3036797;
RX Zehner Z.E., Li Y., Roe B.A., Paterson B.M., Sax C.M.;
RX "The chicken vimentin gene. Nucleotide sequence, regulatory elements,
RX and comparison to the hamster gene.";
RX J. Biol. Chem. 262:8112-8120(1987).
[2]
SEQUENCE OF 453-459 FROM N.A.
RX MEDLINE=83195052; Pubmed=6573660;
RX Zehner Z.E., Paterson B.M.;
RX "Characterization of the chicken vimentin gene: single copy gene
RX producing multiple mRNAs.";
RX Proc. Natl. Acad. Sci. U.S.A. 80:911-915(1983).
-1- FUNCTION: Vimentins are class-III intermediate filaments found in
-1- various non-epithelial cells, especially mesenchymal cells.
-1- SUBUNIT: HOMOPOLYMER.
-1- PTM: ONE OF THE MOST PROMINENT PHOSPHOPROTEINS IN VARIOUS CELLS OF
-1- MESENCHYMAL ORIGIN. PHOSPHORYLATION IS ENHANCED DURING CELL
-1- DIVISION, AT WHICH TIME VIMENTIN FILAMENTS ARE SIGNIFICANTLY
-1- REORGANIZED.
----- BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC

```

RA Dewhurst S., Embretson J.E., Anderson D.C., Mullins J.I., Fultz P.N.;  
 RT Sequence analysis and acute pathogenicity of molecularly cloned  
 RT SIVSM-PB14.  
 RL Nature 345:636-640(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92368737; PubMed=1503826;  
 RA Dewhurst S., Embretson J.E., Fultz P.N., Mullins J.I.;  
 RT Molecular clones from a non-acutely pathogenic derivative of  
 RT SIVsmPB14: characterization and comparison to acutely pathogenic  
 RT clones.  
 RL AIDS Res. Hum. Retroviruses 8:1179-1187(1992).  
 CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage to 5'-  
 CC Phosphomonoester.  
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate  
 CC + {DNA}(N).  
 CC -1- PPM: CLEAVAGE SITES THAT YIELD THE MATURE PROTEINS REMAIN TO BE  
 CC DETERMINED.  
 CC -1- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY A2; ALSO  
 CC KNOWN AS THE RETROPEPSIN FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M31325; AAA47753.1; -  
 DR EMBL: L03298; AAA4777.1; -  
 DR HSSP: P04584; 1PHV.  
 DR HIV: M31325; POLSMPBD.  
 DR MEROPS: A02.002; -  
 DR InterPro: IPR001995; Asparticase\_rtrv.  
 DR InterPro: IPR001969; Asparticase\_sile.  
 DR InterPro: IPR001037; Integrase\_C.  
 DR InterPro: IPR003308; Integrase\_Zn.  
 DR InterPro: IPR002156; RNaseH.  
 DR InterPro: IPR000477; RVTse.  
 DR InterPro: IPR001584; RVTse.  
 DR Pfam: PF00075; rnaaseh.1.  
 DR Pfam: PF00077; rvp.1.  
 DR Pfam: PF00078; rvt.1.  
 DR Pfam: PF00552; Integrase.1.  
 DR Pfam: PF00665; rve.1.  
 DR Pfam: PF02022; Integrase\_Zn.1.  
 DR PROSITE: PS00141; ASP\_PROTEASE.1.  
 DR PROSITE: PS50175; ASP\_PROT\_RETROV.1.  
 KM AIDS: Polyprotein; Hydrolyase; Aspartyl protease; Endonuclease;  
 Nuclease; transferase; RNA-directed DNA polymerase.  
 CHAIN 1 170  
 ACT\_SITE 96 BY SIMILARITY.  
 SEQUENCE 1022 AA; 115869 MW; 16DFBEA03F289D6A CRC64;  
 Query Match 5.18; Score 90.5; DB 1; Length 1022;  
 Best Local Similarity 23.08; Pred. No. 7.6;  
 Matches 56; Conservative 32; Mismatches 92; Indels 63; Gaps 11;  
 QY 16 PRKFGVAGRSCEVLEKRGKRLPERSGRLCYEDTELETDFPSPVNAE---LVL 71  
 DB 265 PHPAGLAKRRIRIVLDGAVFSIP-----LDEERQRTATLTLPV-NNAEGRKRYI 316  
 QY 72 LITGQAWOG---YSDIRFLSAFHE--POVGLIQAAQQL--CDEQAPORQL--LA 120  
 DB 317 KVLPGQKGSPPAIFQHMTRAVLPEPRKANDVTLIQMDIILASDRTDEHDRVVLQK 376  
 QY 121 DLLHNTSNTAATRAEDPPW-FEGLE-----SRFQSGSLRY 158  
 DB 377 ELNLSIGFSPEEKFOEDPPQNGYELMPTKKRLQRIELPQRTWTVDNIQKLVGLNW 436  
 QY 159 SCE-----SRIRSYLRVSSVPTVGAEQEEFLRVGSMQGRRLRSMQYNGSYF 207

DB 437 AAQIYPIKTKHLCRLIRGKMTLEEVQWTEMAEATYEENKI-----ILSQEGECY 489  
 QY 208 DRG 210  
 DB 490 QEG 492  
 RESULT 17  
 DBS\_MOUSE STANDARD; PRT; 1149 AA.  
 AC 064096;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Guanine nucleotide exchange factor DBS (DBL's big sister) (MCF2  
 DE transforming sequence-like protein).  
 GN MCF2L OR DBS.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Hematopoietic;  
 RX MEDLINE=95166559; PubMed=7862449;  
 RA Whitehead I., Kirk H., Kay R.;  
 RT DBS, a homolog of the Dbl guanine nucleotide exchange factor.  
 RL Oncogene 10:713-721(1995).  
 CC -1- FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR THAT POTENTIALLY  
 CC LINKS PATHWAYS THAT SIGNAL THROUGH RAC1, RHOA AND CDC42. CATALYZES  
 CC GUANINE NUCLEOTIDE EXCHANGE ON RHOA AND CDC42 AND INTERACTS  
 CC SPECIFICALLY WITH THE GTP-BOUND FORM OF RAC1, SUGGESTING THAT IT  
 CC FUNCTIONS AS AN EFFECTOR OF RAC1. MAY ALSO PARTICIPATE IN AXONAL  
 CC TRANSPORT IN THE BRAIN. BECOMES ACTIVATED AND HIGHLY TUMORIGENIC  
 CC BY TRUNCATION OF THE N-TERMINUS (BY SIMILARITY)  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LOW LEVELS IN SEVERAL HEMOPOIETIC  
 CC CELL LINES AND IN THYMUS AND SPLEEN, AND AT HIGHER LEVELS IN OTHER  
 CC TISSUES, PARTICULARLY IN BRAIN.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY (DH) DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 CRAL-TRIO DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: S76838; AAB33461.2; -  
 DR MGD: MGI:103263; Mcf2l.  
 DR InterPro: IPR001251; CRAL\_TRIO.  
 DR InterPro: IPR001331; GDS\_CDC24.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RHOGEF.  
 DR InterPro: IPR001452; SH3.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00018; SH3.1.  
 DR Pfam: PF00169; PH.1.  
 DR Pfam: PF00435; spectrin.1.  
 DR Pfam: PF00621; RHOGEF.1.  
 DR SMART: SM00233; PH.1.  
 DR SMART: SM00325; RHOGEF.1.  
 DR SMART: SM00516; SEC14.1.  
 DR SMART: SM00326; SH3.1.  
 DR SMART: SM00150; SPEC.1.  
 DR PROSITE: PS50191; CRAL\_TRIO.1.  
 DR PROSITE: PS00741; DH.1.1.  
 DR PROSITE: PS50010; DH.2.1.  
 DR PROSITE: PS50003; PH\_DOMAIN.1.

DR PROSITE; PS50002; SH3; 1.  
 KW Guanine-nucleotide releasing factor; SH3 domain; Proto-oncogene.  
 FT DOMAIN 52 224 CRAL-TRIO.  
 FT DOMAIN 624 627 POLY-GLU.  
 FT DOMAIN 632 812 DH.  
 FT DOMAIN 841 950 PH.  
 FT DOMAIN 1058 1112 SH3.  
 SQ SEQUENCE 1149 AA; 129148 MW; D910DD133BEA9E3 CRC64;  
 Query Match 5.1%; Score 90.5; DB 1; Length 1149;  
 Best Local Similarity 22.0%; Pred No. 8.8;  
 Matches 74; Conservative 41; Mismatches 134; Indels 87; Gaps 15;  
 47 CLYEDGTELT-EDYFSPVDNAELVLTGLGAGQGYSDIRRLSLAFHEPQYGLQAAQ 105  
 486 CQSDGAEAFSEFQIEKELTGAENKIOELNIEVEYCIINQ-----DLEHYQK 535  
 106 LACDEAPOROLLDLH---NVSQNAIETRAEDP--PWEGL-----E 146  
 536 VF-----OKQESTEMFRRQASLKLAQKOTVPVPAPREALTKSPSPSGWSS 589  
 147 SRFQSKGYLRYSCESIRSYLREY-SSYPSTGAEAEQEFLLVLSMCQRLSMQYNGS 205  
 590 ENSSEGNALRGPRYRAKSEMEPRQRTSTGEE--EESLAI-----LRHVMEL 640  
 206 YFDRGAKGSRCTCPGWFSCGPFDMQSCSRHSINPYNSRESRLIFSTWMLDHIKK 265  
 641 LDFERAYVEELQVLEGYNA-----EMDNPLMALISTGLQNNKNILFG--NMETIYHP 693  
 266 RTIPLTYEA---IKQDGR-----EVDMEFYGLFTSENK 300  
 694 NRNTPAVESCIDCPVLGYRCLEHMEFQYIEKCCNKRRESLIRQSCDPCFQF--- 750  
 301 LVHYCHKTKTKLNDDPSRIYKPTQTRKQPPRK 336  
 751 -----CKKLIDKRLSD-SYLRKPVQRTIKYQLLKK 780  
 RESULT 18  
 AT06\_SCHPO STANDARD; PRT; 1211 AA.  
 ID AT06\_SCHPO Q9UT01; Q9UTD0;  
 AC 014072; Q9UT01; Q9UTD0;  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Probable cation-transporting ATPase C2E11.07C (EC 3.6.3.-).  
 GN SPAC2E11.07C OR SPACUNK4.07C OR SPAPYUK71.01.  
 Schizosaccharomyces pombe (Pilsion Yeast).  
 Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 Schizosaccharomycetales; Schizosaccharomycetaceae;  
 Schizosaccharomyces.  
 NCBI\_TaxID=4896;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 Wood V., Williams R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 Sgouras J., Peat N., Hayes J., Baker S., Basham D., Bowman S.,  
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 Holtroyd S., Hornsby T., Howard S., Huckle E.J., Hunt S., Jagels K.,  
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 Mooney P., Moutle S., Mungall K., Murphy L., O'Donnell D., O'Neill C.,  
 Oliver K., O'Neill S., Pearson S., Saunders D., Seeger K., Sharp S.,  
 Rutherford K., Rutter S., Saunders D., Quail M.A., Rabinowitsch E.,  
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 Woodward J., Voicakeet G., Aert R., Robben J., Gymnopre B.,  
 Weltjens J., Vansteels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 Gabel C., Fuchs M., Fritzc A., Holzer C., Moestl D., Hilbert H.,  
 Borzym K., Langer I., Beck C., Lehnach H., Reinhardt R., Pohl T.M.,  
 Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,

RA goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,  
 RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rocher M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forbush S.L.,  
 RA Cerutti L., Lowe T., McCombe W.R., Paulsen I., Potashkin J.,  
 RA Shporkovskii G.V., Usery D., Barrell B.G., Nurse P.,  
 RA "The genome sequence of Schizosaccharomyces pombe";  
 RT Nature 415:871-880(2002).  
 RN [2]  
 RN SEQUENCE OF 617-809 FROM N.A., AND SUBCELLULAR LOCATION.  
 RC STRAIN=968 h50;  
 RC MEDLINE=20223868; PubMed=10759889; Chikashige Y., Harauchi T.,  
 RX Ding D.Q., Tomita Y., Yamamoto A., Chikashige Y., Harauchi T.,  
 RA Hirooka Y.;  
 RT "Large-scale screening of intracellular protein localization in living  
 RT fission yeast cells by the use of a GFP-fusion genomic DNA library";  
 RT genes Cells 5:169-190(2000).  
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2  
 CC ATPASES). SUBFAMILY V.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.slb.ch/announce/>  
 CC or send an email to [license@isb.slb.ch](mailto:license@isb.slb.ch)).  
 CC -----  
 DR EMBL; AL031190; CAN20137.1; -  
 DR EMBL; AL037738; CAB22144.1; -  
 DR EMBL; AB027853; BAA87157.1; -  
 DR InterPro; IPR001757; ATPase\_E1-E2.  
 DR InterPro; IPR001454; Hlgase/Hydrlase.  
 DR Pfam; PF00122; E1-E2\_ATPase; 1.  
 DR Pfam; PF00702; Hydrolase; 1.  
 DR PRINTS; PR00119; CATATPASE.  
 DR PROSITE; PS00154; ATPASE\_E1\_E2; 1.  
 DR HYDROLASE; Transmembrane; Phosphorylation; Magnesium; ATP-binding.  
 KW DOMAIN 1 21  
 FT TRANSSEM 22 43  
 FT DOMAIN 44 49  
 FT TRANSSEM 50 72  
 FT DOMAIN 73 191  
 FT TRANSSEM 192 214  
 FT DOMAIN 215 217  
 FT TRANSSEM 218 236  
 FT DOMAIN 237 397  
 FT TRANSSEM 398 417  
 FT DOMAIN 418 430  
 FT TRANSSEM 431 452  
 FT DOMAIN 453 495  
 FT TRANSSEM 496 1015  
 FT DOMAIN 1016 1022  
 FT TRANSSEM 1023 1039  
 FT DOMAIN 1040 1057  
 FT TRANSSEM 1058 1081  
 FT DOMAIN 1082 1101  
 FT TRANSSEM 1102 1124  
 FT DOMAIN 1125 1135  
 FT TRANSSEM 1136 1155  
 FT DOMAIN 1156 1172  
 FT TRANSSEM 1173 1195  
 FT DOMAIN 1196 1211  
 FT MOD\_RES 485 485  
 FT METAL 824 824  
 FT METAL 828 828  
 SQ SEQUENCE 1211 AA; 136260 MW; 6EEB228CA5A57EC CRC64;  
 Query Match 5.0%; Score 90; DB 1; Length 1211;  
 Best Local Similarity 20.7%; Pred. No. 10;

Matches 64; Conservative 42; Mismatches 111; Indels 92; Gaps 12;

QY 37 FOLPERGS-RLCTEDTELTEDYPPSPVDAEVLTLGAMOGYSDIR-----FL 89  
 Db 471 FRILSGHLIDICCDKCTGLTEEH-----WVGIGAGVNRKRDYSLEKL 514  
 QY 90 SAFHEPOYGLIQAAOQLCDEAPORORLAD-----LHVNSONIAETAEDEP--W 141  
 Db 515 SDASNDALATATATHTLVLEBGETPRVGVDPPEKATVENLGSITEKNFVSAEGSVF 574  
 QY 142 FEG---LESRPQSGYLRYSCESRIRSYLREVSYPSTVGAEOEFLRYLGSQCRIR 198  
 Db 575 YKGVQIIRNFQFSALKRQSSVSNVRSGSFTEFVSQAP-----EVIATMLREV- 627  
 QY 199 SMQNGSFPRGAGGSRILCPPEGFCQCPDMDCLSRHSINPYNSRESRIILFTW-- 256  
 Db 628 PRDEKITYKDGKRGSRVALGKRYFKNYIPENOVSDLSRESI-----ESDLVFAGFLI 681  
 257 -----NLDH-----ITEKRTIIPITVEAIKE 278  
 682 FTSPLEKEDARQTVOMLNNSHRCAMITGDNPLTAVYAEQVGIYER-----PTLVLDIKH 736  
 279 QDGEVDWE 287  
 737 ENKILEWK 745

RESULT 19  
 G6PE\_RABIT  
 ID G6PE\_RABIT STANDARD; PRT; 763 AA.  
 AC P56201;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE G6P/66GL endoplasmic b1functional protein [includes: Glucose 1-dehydrogenase (EC 1.1.1.47) (Hexose-6-phosphate dehydrogenase); 6-phosphoglucanone-lactonase (EC 3.1.1.31) (6PGL)].  
 GN G6P.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE.  
 RC SPRAIN-New Zealand white; TISSUE=Liver;  
 RX MEDLINE=93281746; PubMed=8506377;  
 RA "Isolation and the complete amino acid sequence of luminal endoplasmic reticulum glucose-6-phosphate dehydrogenase.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 90:5302-5306(1993)  
 RL  
 -1- FUNCTION: OXIDIZES GLUCOSE-6-PHOSPHATE AND GLUCOSE, AS WELL AS OTHER HEXOSE-6-PHOSPHATES (BY SIMILARITY).  
 -1- CATALYTIC ACTIVITY: Beta-D-glucose + NAD(P)(+) = D-glucono-1,5-lactone + NAD(P)H.  
 -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NAD(P)(+) = D-glucono-1,5-lactone 6-phosphate + NAD(P)H.  
 -1- CATALYTIC ACTIVITY: 6-phospho-D-glucono-1,5-lactone + H(2)O = 6-phospho-D-gluconate.  
 -1- COFACTOR: CAN USE BOTH NAD OR NADP.  
 -1- SUBCELLULAR LOCATION: MICROSOSE, ENDOPLASMIC RETICULUM LUMEN.  
 -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE FAMILY.  
 -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GLUCOSAMINE/GALACTOSAMINE-6-PHOSPHATE ISOMERASE FAMILY.  
 -1- CAUTION: THE SEQUENCE IN REF.1 SEEMS INCORRECT DUE TO THE TRANSPOSITION OF A SECTION THAT WAS PUT IN POSITION 330 TO 436 AND WHICH IS NOW TRANSPOSED TO POSITION 593 TO 700.  
 DR InterPro: IPR001282; G6PD.  
 DR InterPro: IPR000457; Glucosamine\_iso.  
 DR Pfam: PF00479; G6PD; 1.  
 DR Pfam: PF01182; Glucosamine\_iso; 1.  
 DR Pfam: PF02781; G6PD\_C; 1.

DR PRINTS: PR00079; G6PDHGRNASE.  
 DR PRODOM: PD001129; G6PD; 1.  
 DR TIGRFAMs: TIGR01198; pgl; 1.  
 DR PROSITE: PS00069; G6P\_DEHYDROGENASE; 1.  
 KW Oxidoreductase; NAD; NADP; Glucose metabolism; Endoplasmic reticulum;  
 KW Glycoprotein; Hydrolase; Multifunctional enzyme.  
 FT MOD\_RES 1  
 FT DOMAIN 1 507  
 FT DOMAIN 508 521  
 FT DOMAIN 522 573  
 FT ACT\_SITE 189 189  
 FT CARBOHYD 138 138  
 FT CARBOHYD 263 263  
 FT VARIANT 69  
 FT VARIANT 75  
 SO SEQUENCE 763 AA; 85285 MW; 2A4E10390628B672 CRC64;

Query Match 5.0%; Score 89.5; DB 1; Length 763;

Best Local Similarity 23.7%; Pred. No. 6.3; Mismatches 110; Indels 85; Gaps 18;

Matches 75; Conservative 47; Mismatches 110; Indels 85; Gaps 18;

QY 76 QAMOGYSDIRRFSLAFHE-POVGLIQAAOQLCDE-----QAP-QORRL 119  
 Db 1 QELGHSVY--LIGATGDLAKTYLMQGLFPLFDEAGKSHSFPHGALTAAPQGGELM 58  
 QY 120 ADLHANS--ONIAETAEDEPPWFEGLSEKFSKSGYLRYSCESRIRSYLREVSSTPST 177  
 Db 59 AKALESLSQPRDMAPSLCAE-----LQAFRLRSRYRLQKTAEDYQALGRDIE----- 106  
 QY 178 VGAEOEFLRVLSM-----CORLSMQNG-----SYFPGAKGG 214  
 Db 107 --AOYQDGLREAGMFFTSPPRAYDIANNINSSCPGAGMLRVYLEPFGHDLSA 164  
 QY 215 SRLCTPEG-WFSCGCPEDMDCLSRHS--INPYNSRESRIILFTWMDHITTEKRTIP 270  
 Db 165 QQLATELGSFQOEEMRYVDHLYGKQAVQILPRDQRRALDSLMNRHH-VEVEIIMK 223  
 QY 271 TLVEAIKQDQREVDMEFFGLL-----FISENLKY-----HYVHKRT-THKLCD 317  
 Db 224 ETVDA---EGRTSFYER-YGVINDTLQNLHTEITLVAMELPANVSCSEAVLRHKL--- 275  
 QY 318 PSRIYKQTRLRKROPV 334  
 Db 276 --QAFRALRRLRQGSAY 290

RESULT 20  
 DESM\_XENLA  
 ID DESM\_XENLA STANDARD; PRT; 458 AA.  
 AC P23239;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-NOV-1991 (Rel. 20, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Desmnl.  
 GN DES.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90032404; PubMed=2806128;  
 RA Herrmann H., Fouquet B., Franke W.W.;  
 RA "Expression of intermediate filament proteins during development of Xenopus laevis. II. Identification and molecular characterization of desmin.";  
 RT Development 105:299-307(1989).  
 RL  
 -1- FUNCTION: DESMIN ARE CLASS-III INTERMEDIATE FILAMENTS FOUND IN MUSCLE CELLS. IN ADULT STRIATED MUSCLE THEY FORM A FIBROUS NETWORK CONNECTING MYOFIBRILS TO EACH OTHER AND TO THE PLASMA MEMBRANE FROM THE PERIPHERY OF THE Z-LINE STRUCTURES.  
 -1- SUBUNIT: HOMOPOLYMER.

CC -! SUBCELLULAR LOCATION: Cytoplasmic.  
CC -! SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed, usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; X16842; CA34740.1; -  
DR PIR; A43534; A43534.  
DR InterPro; IPR001664; IF.  
DR Pfam; PF00038; Filament; 1.  
DR PROSITE; PS00226; IF; 1.  
DR Intermediate filament; Coiled coil; Muscle protein.  
DOMAIN 1 96 HEAD.  
DOMAIN 97 400 ROD.  
DOMAIN 401 458 TAIL.  
FT DOMAIN 97 129 COIL 1A.  
FT DOMAIN 130 139 LINKER 1.  
FT DOMAIN 140 240 COIL 1B.  
FT DOMAIN 241 256 LINKER 12.  
FT DOMAIN 257 275 COIL 2A.  
FT DOMAIN 276 283 LINKER 2.  
FT DOMAIN 284 400 COIL 2B.  
FT SITE 342 342 STUTTER.  
SQ SEQUENCE 458 AA; 52802 MW; E96167E1D8FDEB74 CRC64;  
  
Query Match 4.9%; Score 88.5; DB 1; Length 458;  
Best local similarity 24.3%; Pred. No. 4;  
Matches 49; Conservative 24; Mismatches 72; Indels 57; Gaps 10;  
  
QY 50 EDGHELED-----YFSPVPMALVILITGAMOGYVSDIRPLSAFHEPOVGLIOA- 102  
DB 183 QEEIQLEKDENNLAARFGVDATLARIIDLERKIESLQDET-AFLKTHIEIRELQAO 241  
QY 103 --AQQL--CDEAPQORLLADLHNVSQNIATRAEDPPWF----- 142  
DB 242 FQEQQLQVEIDVSKPDLTALDIDRAQY-ENIAAKNVAEAEEWYKSKVSDLNQAAKKNND 300  
QY 143 -----EGLESFQSKSGLYRSC-----SRIR-----SYLREVSYPSTVG 179  
DB 301 AMRSKQEMMEYRHQIOS-----YTCEIDALKGTNDSIMROMDLEERKFGSAAGYODTIG 356  
QY 180 AEAQEEFLRVIGSMCORLSMQ 201  
357 -RLEEIRNNKDEMAHRLREYQ 377

Search completed: May 26, 2003, 15:21:38  
Job time : 23 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd

OM protein - protein search, using sw model

Run on: May 26, 2003, 12:08:32 ; Search time 55 Seconds  
(without alignments)  
818.886 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1789  
Sequence: 1 MLDKPKSVKLRLRSPKFG.....SRIYQTRLKRKQPVRRKQ 338

ing table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Maximum DB seq length: 200000000000

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

1	/SID52/gcgdata/genseq/genseqp-emb1/AA1980.DAT *
2	/SID52/gcgdata/genseq/genseqp-emb1/AA1981.DAT *
3	/SID52/gcgdata/genseq/genseqp-emb1/AA1982.DAT *
4	/SID52/gcgdata/genseq/genseqp-emb1/AA1983.DAT *
5	/SID52/gcgdata/genseq/genseqp-emb1/AA1984.DAT *
6	/SID52/gcgdata/genseq/genseqp-emb1/AA1985.DAT *
7	/SID52/gcgdata/genseq/genseqp-emb1/AA1986.DAT *
8	/SID52/gcgdata/genseq/genseqp-emb1/AA1987.DAT *
9	/SID52/gcgdata/genseq/genseqp-emb1/AA1988.DAT *
10	/SID52/gcgdata/genseq/genseqp-emb1/AA1989.DAT *
11	/SID52/gcgdata/genseq/genseqp-emb1/AA1990.DAT *
12	/SID52/gcgdata/genseq/genseqp-emb1/AA1991.DAT *
13	/SID52/gcgdata/genseq/genseqp-emb1/AA1992.DAT *
14	/SID52/gcgdata/genseq/genseqp-emb1/AA1993.DAT *
15	/SID52/gcgdata/genseq/genseqp-emb1/AA1994.DAT *
16	/SID52/gcgdata/genseq/genseqp-emb1/AA1995.DAT *
17	/SID52/gcgdata/genseq/genseqp-emb1/AA1996.DAT *
18	/SID52/gcgdata/genseq/genseqp-emb1/AA1997.DAT *
19	/SID52/gcgdata/genseq/genseqp-emb1/AA1998.DAT *
20	/SID52/gcgdata/genseq/genseqp-emb1/AA1999.DAT *
21	/SID52/gcgdata/genseq/genseqp-emb1/AA2000.DAT *
22	/SID52/gcgdata/genseq/genseqp-emb1/AA2001.DAT *
23	/SID52/gcgdata/genseq/genseqp-emb1/AA2002.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1789	100.0	338	21	AA57440	Human DNA fragment
2	1789	100.0	338	22	AA663592	Amino acid sequenc
3	1789	100.0	338	22	AA72467	Human DNA fragment
4	1786	99.8	338	22	AA74123	Caspase activated
5	1416	79.2	344	20	AA72931	Mouse caspase-acti
6	1416	79.2	344	20	AA72936	Mouse caspase acti
7	372	20.8	96	22	AA009550	Human polypeptide
8	321	17.9	450	22	AB864406	Drosophila melanog
9	99	5.5	466	21	AA793335	Human vimentin. H
10	99	5.5	466	22	AA866349	Murine vimentin.

1	1.1	99	4.66	23	ABB5731.0	
2	1.1	99	5.5	14	AAW5435.1	
3	1.3	97	5.4	466	21	AAB29635.3
4	1.4	97	5.4	466	22	AAB29635.48
5	1.5	97	5.5	466	23	AAU876837.8
6	1.6	94.5	5.3	1152	22	AAW79837.3
7	1.7	93.5	5.2	1120	22	AAW7885.1
8	1.8	93	5.2	181	22	AAW40801.5
9	1.9	93	5.2	214	22	AAU03204.0
10	2.0	92.5	5.2	221	22	ABG04580.2
11	2.1	92.5	5.2	238	21	AAW94441.1
12	2.2	92.5	5.2	238	21	AAW57607.0
13	2.3	92.5	5.2	238	22	AAW40366.2
14	2.4	92.5	5.2	238	22	ABB63825.5
15	2.5	91.5	5.1	907	20	AAW34898.8
16	2.6	91	5.1	219	21	AAW5841.0
17	2.7	91	5.1	1719	22	ABB70837.3
18	2.8	89.4	5.0	5392	22	ABB65422.2
19	2.9	88	5.0	1135	20	AAW42135.2
20	3.0	88	5.0	142	22	ABG2871.5
21	3.1	88	5.0	160	22	AAW42135.2
22	3.2	89	5.0	331	21	AAW51744.4
23	3.3	89	5.0	331	22	AAW65595.9
24	3.4	88.5	4.9	331	22	AAW774456.8
25	3.5	88.5	4.9	429	23	ABP47758.5
26	3.6	88.5	4.9	465	23	ABB97339.8
27	3.7	88.5	4.9	380	10	AAW37298.9
28	3.8	88.5	4.9	3210	9	AAW81771.1
29	3.9	88	4.9	613	22	AAW81771.1
30	4.0	88	4.9	756	21	AAW5927.7
31	4.1	87.5	4.9	666	22	ABB67103.7
32	4.2	87.5	4.9	1035	11	AAW40522.3
33	4.3	87	4.9	331	19	AAW04012.1
34	4.4	87	4.9	331	21	AAW04192.1
35	4.5	87	4.9	366	21	AAW5691.21

## ALIGNMENTS

RESULT 1  
AAV57440  
ID AAV57440 standard; Protein; 338 AA.  
AC AAV57440;  
XX  
XX  
XX 22-FEB-2000 (first entry)  
DT  
DE Human DNA fragmentation factor DFF40 protein sequence.  
DE  
KW Human; DNA fragmentation factor; DFF40; DFF45; apoptosis; gene therapy  
XX cytosolic; growth; tumour.  
XX  
XX OS Homo sapiens.  
OS  
PN MO954482-A1.  
PN  
PD 28-OCT-1999.  
PD  
PF 16-APR-1998; 98WO-US07895.  
XX  
XX 16-APR-1998; 98WO-US07895.  
PR  
XX  
PA (TEXA ) UNIV TEXAS SYSTEM.  
PA  
PI Wang X, Liu X;  
PI  
DR WPI: 2000-052702/04.  
DR N-PDB: AAZ8592.  
XX  
XX DNA fragmentation factor DFF40 involved in apoptosis and related  
PT polynucleotide -  
XX

PS Claim 2; Page 127-128; 154pp; English.  
 XX  
 CC The present sequence represents a human DNA fragmentation factor,  
 CC designated DFF40. Also described are: (1) a method of inducing apoptosis  
 CC in a cell comprising providing the cell with DFF40 which results in  
 CC apoptosis; (2) a method for inhibiting the growth of a cancer cell  
 CC comprising contacting a cancer cell with a DNA fragmentation factor  
 CC designated DFF40 under conditions permitting the uptake of the DNA  
 CC fragmentation factor by the cell where the presence of the DFF40 into  
 CC the cell induces apoptosis; (3) a method for treating cancer comprising:  
 CC (a) encoding a DFF40 DNA fragmentation factor; and (b) a promoter active  
 CC in the tumour cell, where the promoter is operably linked to the region  
 CC encoding the DNA fragmentation factor, under conditions permitting the  
 CC uptake of the nucleic acid by the tumour cell; (4) a method of  
 CC identifying a modulator of DFF40; and (5) a method of producing a  
 CC functional DNA fragmentation factor. An expression construct encoding a  
 CC DNA fragmentation factor DFF40 and DFF45 complex is provided to a cell to  
 CC induce apoptosis, especially in tumour cells. DFF40 is used to inhibit  
 CC the growth of a cancer cell, especially in humans.

XX Sequence 338 AA;

Query Match 100.0%; Score 1789; DB 21; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 8,8e-167;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLOPKSVKTLALSPKRFVAGRSCEVLRKGCRLRQLPBERGSRCLYEDGETELTDEYF 60  
 Db 1 MLOPKSVKTLALSPKRFVAGRSCEVLRKGCRLRQLPBERGSRCLYEDGETELTDEYF 60  
 QY 61 PSVDNMEVLTLTGAMOGVSDIRRFLSAFHEPQVGLIOAAOQLCDEAPQRRLA 120  
 Db 61 PSVDNMEVLTLTGAMOGVSDIRRFLSAFHEPQVGLIOAAOQLCDEAPQRRLA 120  
 QY 121 DLHNVSONIAETRAEDPWFEGLESRFQSGYLRYSCESRIRSYLREVSYPSTVGA 180  
 Db 121 DLHNVSONIAETRAEDPWFEGLESRFQSGYLRYSCESRIRSYLREVSYPSTVGA 180  
 QY 181 EAOEFLRVLGSMQORLSMOYNSYFDRGAKGSRCTPEGWSGCGPFDMDSCLSRHS 240  
 Db 181 EAOEFLRVLGSMQORLSMOYNSYFDRGAKGSRCTPEGWSGCGPFDMDSCLSRHS 240  
 QY 241 INPYSNESRILFSTWMDHIEKKRTIPLVAIKEODGREVDMEYFGLLFTSENK 300  
 Db 241 INPYSNESRILFSTWMDHIEKKRTIPLVAIKEODGREVDMEYFGLLFTSENK 300  
 QY 301 LVHIVCHAKTTHKLNCDPSRIYKQTRLKRRQPVRRKQ 338  
 Db 301 LVHIVCHAKTTHKLNCDPSRIYKQTRLKRRQPVRRKQ 338

RESULT 2  
 MAG63592 standard; Protein; 338 AA.

AAAG63592;

DT 15-OCT-2001 (first entry)

DE Amino acid sequence of human DNA fragmentation factor 40 (DFF40).

XX Human: DNA fragmentation factor; DFF40; DFF45; apoptosis; DNase;

KW molecular chaperone; cancer cell.

XX Homo sapiens.

XX US2001011078-A1.

XX 02-AUG-2001.

XX 22-DEC-2000; 2000US-0748451.

XX 16-APR-1998; 98US-0065742.

XX (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 XX Wang X, Liu X;  
 XX WPI: 2001-496169/54.  
 DR N-PSDB; AAH74642.

PT New DNA fragmentation factor polypeptides and polynucleotides, useful  
 PT for inhibiting the growth of cancer cells, as well as for inducing  
 PT apoptosis of cells.

PS Claim 2; Page 45-46; 56pp; English.

XX The present sequence represents a human DNA fragmentation factor subunit  
 CC of 40 kDa, designated DFF40. The specification also describes DFF45.  
 CC DFF40 is capable of inducing apoptosis, and may contain a nuclear  
 CC localisation fragment. DFF45 acts as a molecular chaperone to direct  
 CC the folding of DFF40. Although all DNase activity is associated with  
 CC DFF40, DFF activity only occurs once DFF40 is complexed with DFF45.  
 CC The DFF polypeptides and polynucleotides are useful for inhibiting  
 CC the growth of cancer cells, and for inducing apoptosis of cells.

SQ Sequence 338 AA;

Query Match 100.0%; Score 1789; DB 22; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 8,8e-167;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLOPKSVKTLALSPKRFVAGRSCEVLRKGCRLRQLPBERGSRCLYEDGETELTDEYF 60  
 Db 1 MLOPKSVKTLALSPKRFVAGRSCEVLRKGCRLRQLPBERGSRCLYEDGETELTDEYF 60  
 QY 61 PSVDNMEVLTLTGAMOGVSDIRRFLSAFHEPQVGLIOAAOQLCDEAPQRRLA 120  
 Db 61 PSVDNMEVLTLTGAMOGVSDIRRFLSAFHEPQVGLIOAAOQLCDEAPQRRLA 120  
 QY 121 DLHNVSONIAETRAEDPWFEGLESRFQSGYLRYSCESRIRSYLREVSYPSTVGA 180  
 Db 121 DLHNVSONIAETRAEDPWFEGLESRFQSGYLRYSCESRIRSYLREVSYPSTVGA 180  
 QY 181 EAOEFLRVLGSMQORLSMOYNSYFDRGAKGSRCTPEGWSGCGPFDMDSCLSRHS 240  
 Db 181 EAOEFLRVLGSMQORLSMOYNSYFDRGAKGSRCTPEGWSGCGPFDMDSCLSRHS 240  
 QY 241 INPYSNESRILFSTWMDHIEKKRTIPLVAIKEODGREVDMEYFGLLFTSENK 300  
 Db 241 INPYSNESRILFSTWMDHIEKKRTIPLVAIKEODGREVDMEYFGLLFTSENK 300  
 QY 301 LVHIVCHAKTTHKLNCDPSRIYKQTRLKRRQPVRRKQ 338  
 Db 301 LVHIVCHAKTTHKLNCDPSRIYKQTRLKRRQPVRRKQ 338

RESULT 3

AAI72467 standard; Protein; 338 AA.

AAI72467;

DT 24-APR-2001 (first entry)

DE Human DNA fragmentation factor 40 (DFF40).

XX Human: DNA fragmentation factor; DFF; apoptosis; molecular chaperone;

KW gene therapy; hyperproliferative disorder; therapy; tumour; restenosis;

XX psoriasis; angiogenesis; cancer; cytosolic; neoplasia.

XX Homo sapiens.

XX US6165737-A.

XX 26-DEC-2000.

XX 16-APR-1998; 98US-0061702.  
 XX 16-APR-1998; 98US-0061702.  
 PR 16-APR-1998; 98US-0061702.

PA (TEXA ) UNIV TEXAS SYSTEM.

PI Wang X, Liu X;

DR WPI: 2001-090481/10.

DR N-PSDB: AAD02498.

XX Identifying modulator of human DNA fragmentation factor 40, for  
 PT treating cancer, involves contacting cell or cell-free composition  
 PR comprising DFF40 with candidate substance and comparing apoptosis with  
 control

Example 1; Fig 1; 52pp; English.

XX The present sequence is a human DNA fragmentation factor 40 (DFF40)  
 CC capable of inducing apoptosis. DFF is a heterodimeric protein  
 CC comprising 40kDa and 45kDa subunits. DFF45 acts as a molecular chaperone  
 CC to facilitate the appropriate folding of DFF40 and acts as an inhibitor  
 CC for DFF40. DFF40 and DFF45 are used in gene therapy. The modulators of  
 CC human DFF40 activity are useful for inducing apoptosis and for treating  
 CC hyperproliferative disorders such as restenosis, psoriasis, metastatic  
 CC tumours, angiogenesis and benign and malignant neoplasias. They are also  
 CC used for treating cancers of the brain (glioblastoma, astrocytoma,  
 CC oligodendroglioma and ependymoma), lung (liver, spleen, kidney, lymph  
 CC node, pancreas, small intestine, blood cells, colon, stomach, breast,  
 CC endometrium, prostate, testicle, ovary, skin, head and neck, oesophagus,  
 CC bone marrow, blood, other tissue and multi-drug resistant cancer.

XX Sequence 338 AA;

Query Match 100.0%; Score 1789; DB 22; Length 338;  
 Best Local Similarity 100.0%; Pred. No. 8, 8e-167;  
 Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLOPKSVKLRALRSPKRGVAGRSQCEVLRKGLRQLPDRGSRCLTYEDGTETEDYF 60  
 DB 1 MLOPKSVKLRALRSPKRGVAGRSQCEVLRKGLRQLPDRGSRCLTYEDGTETEDYF 60  
 QY 61 PSYDPAEVLITLGOAMOGYSDIRRFISAFHEPOVGLIOAQQLLCDEQAPORORLLA 120  
 DB 61 PSYDPAEVLITLGOAMOGYSDIRRFISAFHEPOVGLIOAQQLLCDEQAPORORLLA 120  
 QY 121 DLHNVSONIAETRAEDPPMEFEGLESFRQSGYLRSYSCESRIRSYLREVSSTYVGA 180  
 DB 121 DLHNVSONIAETRAEDPPMEFEGLESFRQSGYLRSYSCESRIRSYLREVSSTYVGA 180  
 QY 181 EAOEFLRVLSGMCQRLRSMOYNGSYFDGAGGSRCLCTPGWFSQCGPDMDSCLRHS 240  
 DB 181 EAOEFLRVLSGMCQRLRSMOYNGSYFDGAGGSRCLCTPGWFSQCGPDMDSCLRHS 240  
 QY 241 INPYSNRESRILESTWMLDHIIEKKRTIIPLVLAIEKQDREVDMEYFYGLFTSENK 300  
 DB 241 INPYSNRESRILESTWMLDHIIEKKRTIIPLVLAIEKQDREVDMEYFYGLFTSENK 300  
 QY 301 LVHIVCHKTKTHKLNCDPSRIYKPOTRLKRQPVARRQ 338  
 DB 301 LVHIVCHKTKTHKLNCDPSRIYKPOTRLKRQPVARRQ 338

RESULT 4

AA04123 standard; Protein; 338 AA.

AC AA04123;

DT 11-JUN-1999 (first entry)

DE Caspase activated nuclease CPAN.

XX Caspase activated nuclease; CPAN; apoptosis; human; heart attack;  
 KW stroke; tumour.  
 XX Homo sapiens.  
 OS WO9910501-A1.  
 XX 04-MAR-1999.  
 PD 19-AUG-1998; 98MO-US17214.  
 PF 06-JAN-1998; 98US-0072192.  
 PR 22-AUG-1997; 97US-0056904.  
 PR 22-AUG-1997; 97US-0056907.

XX (CHIR ) CHIRON CORP.

PI Halenbeck R, MacDonald H;

DR WPI: 1999-190620/16.

DR N-PSDB: AX19907.

XX New isolated caspase activated nuclease - used to develop products  
 PT which can inhibit or activate apoptosis for use in treating e.g.  
 PT heart attack, stroke or tumours  
 PS Claim 1; Page 38; 61pp; English.

XX The present sequence is a caspase activated nuclease (CPAN). Active  
 CC CPAN proteins can be used for identifying compounds which inhibit  
 CC apoptosis. Inhibitors of apoptosis can be used therapeutically, e.g.  
 CC where cells are damaged and the apoptotic pathway is initiated such as  
 CC by heart attack or stroke. The inactive CPAN proteins can be used for  
 CC identifying compounds which activate apoptosis. Compounds which activate  
 CC apoptosis can be used to treat tumours. The association of CPAN with a  
 CC caspase-sensitive inhibitor suggests that when cells enter into  
 CC apoptosis, the activation of caspases will trigger CPAN to become  
 CC active.

XX Sequence 338 AA;

Query Match 99.8%; Score 1786; DB 20; Length 338;  
 Best Local Similarity 99.7%; Pred. No. 1, 7e-166;  
 Matches 337; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLOPKSVKLRALRSPKRGVAGRSQCEVLRKGLRQLPDRGSRCLTYEDGTETEDYF 60  
 DB 1 MLOPKSVKLRALRSPKRGVAGRSQCEVLRKGLRQLPDRGSRCLTYEDGTETEDYF 60  
 QY 61 PSYDPAEVLITLGOAMOGYSDIRRFISAFHEPOVGLIOAQQLLCDEQAPORORLLA 120  
 DB 61 PSYDPAEVLITLGOAMOGYSDIRRFISAFHEPOVGLIOAQQLLCDEQAPORORLLA 120  
 QY 121 DLHNVSONIAETRAEDPPMEFEGLESFRQSGYLRSYSCESRIRSYLREVSSTYVGA 180  
 DB 121 DLHNVSONIAETRAEDPPMEFEGLESFRQSGYLRSYSCESRIRSYLREVSSTYVGA 180  
 QY 181 EAOEFLRVLSGMCQRLRSMOYNGSYFDGAGGSRCLCTPGWFSQCGPDMDSCLRHS 240  
 DB 181 EAOEFLRVLSGMCQRLRSMOYNGSYFDGAGGSRCLCTPGWFSQCGPDMDSCLRHS 240  
 QY 241 INPYSNRESRILESTWMLDHIIEKKRTIIPLVLAIEKQDREVDMEYFYGLFTSENK 300  
 DB 241 INPYSNRESRILESTWMLDHIIEKKRTIIPLVLAIEKQDREVDMEYFYGLFTSENK 300  
 QY 301 LVHIVCHKTKTHKLNCDPSRIYKPOTRLKRQPVARRQ 338  
 DB 301 LVHIVCHKTKTHKLNCDPSRIYKPOTRLKRQPVARRQ 338

RESULT 5

AA029931

ID AAY29931 standard; Protein: 344 AA.  
 AC AAY29931;  
 DT 23-NOV-1999 (first entry)  
 DE Mouse caspase-activating DNase.  
 XX Mouse caspase-activating DNase.  
 XX Mouse; caspase activating DNase; CAD; inhibitor; ICAD; ICAD-L; ICAD-S;  
 KM cell death.  
 OS Mus sp.  
 PN JP11239495-A.  
 PD 07-SEP-1999.  
 PF 25-DEC-1998; 98JP-0369222.  
 PR 25-DEC-1997; 97JP-0369356.  
 (OSAB-) ZH OSAKA BIOSCIENCE KENKYUSHO.

WPI: 1999-554024/47.  
 N-PSDB: AA221173.

Inhibitor for caspase-activating DNase - useful for preventing cell death and is useful as reagent for cell death

Example; Page 16-17; 26pp; Japanese.

The present invention describes an inhibitor (ICAD) for caspase-activating DNase (CAD). Two ICADs are specifically claimed: a long chain form (ICAD-L) and a short chain form (ICAD-S). ICAD and ICAD CDNA can be used in the prevention of cell death, and can be used as a research reagent for cell death. The present sequence represents mouse CAD which is used in the exemplification of the present invention.

Sequence 344 AA;

Query Match 79.2%; Score 1416; DB 20; Length 344;  
 Best Local Similarity 77.2%; Pred. No. 3.6e-130;  
 Matches 261; Conservative 37; Mismatches 38; Indels 2; Gaps 1;

QY 1 MLDKPKSVKRLALNSPKRFVAGNSCOEVLKRGCLRFQLPERSRLCLYEDGTELTEDYF 60  
 DB 4 VLROPKCYKRLALNSACKFVGAARSCQELLRKGCVRFLQPMGSRCLCYEDGTEVTDCCF 63  
 QY 61 PSVDNMLVLTGAGMOGVSDIRRLSFHPEOVGLIOAQQILCDEQAPQORLLA 120  
 DB 64 GLPNDDELTLTAGEWGHVSDITRPLSVFNEPHAGVIOAQQILSDEQAPLROKLLA 123  
 QY 121 DLHVNSONIAETRAEDPPWFEGLESFQSGYLRSYCSRSIRSYLREVSYPSTVGA 180  
 DB 124 DLHVNSONIAETRAEDPPWFEGLESFQSGYLRSYCSRSIRSYLREVSATYSWDE 183  
 QY 181 EAQEFPLAVLSCQRLSMOYNGSYFDRGAKGSRCTPGWFSQGPFDMSCLSRHS 240  
 DB 184 AAOEYFLVLSMCKLKSVOYNGSYFDRGAEASRLCTPGWFSQGPFDMSCLSRHS 243  
 QY 241 INPYSNRSRLFTSTWINDHIEKRTIPLVNAIKQDGRVDMEFYLLTSENLK 300  
 DB 244 INPYSNRSRLFTSTWINDHIEKRTIPLVNAIKQDGRVDMEFYLLTSENLK 301  
 QY 301 LVHIVCHKKTTHKLCNCPDSRIYKQPTRLKRPYARRQ 338  
 DB 302 LVHIVCHKKTTHKLCNCPDSRIYKQPTRLKRPYARRQ 339

RESULT 6  
 AAY29936  
 ID AAY29936 standard; Protein: 344 AA.

AC AAY29936;  
 DT 23-NOV-1999 (first entry)  
 DE Mouse caspase activating DNase.  
 XX Mouse; caspase activating DNase; CAD; inhibitor; ICAD; ICAD-L; ICAD-S;  
 KM cell death; cancer.  
 OS Mus sp.  
 PN JP11239494-A.  
 PD 07-SEP-1999.  
 PF 25-DEC-1998; 98JP-0369093.  
 PR 25-DEC-1997; 97JP-0369443.  
 (OSAB-) ZH OSAKA BIOSCIENCE KENKYUSHO.

WPI: 1999-554023/47.  
 N-PSDB: AA221180.

New caspase-activating DNase - useful for eliminating cancer cells  
 Claim 1; Page 14-15; 20pp; Japanese.

The present sequence represents mouse caspase activating DNase (CAD).  
 CAD can be used for the removal of cancer cells.

Sequence 344 AA;

Query Match 79.2%; Score 1416; DB 20; Length 344;  
 Best Local Similarity 77.2%; Pred. No. 3.6e-130;  
 Matches 261; Conservative 37; Mismatches 38; Indels 2; Gaps 1;

QY 1 MLDKPKSVKRLALNSPKRFVAGNSCOEVLKRGCLRFQLPERSRLCLYEDGTELTEDYF 60  
 DB 4 VLROPKCYKRLALNSACKFVGAARSCQELLRKGCVRFLQPMGSRCLCYEDGTEVTDCCF 63  
 QY 61 PSVDNMLVLTGAGMOGVSDIRRLSFHPEOVGLIOAQQILCDEQAPQORLLA 120  
 DB 64 GLPNDDELTLTAGEWGHVSDITRPLSVFNEPHAGVIOAQQILSDEQAPLROKLLA 123  
 QY 121 DLHVNSONIAETRAEDPPWFEGLESFQSGYLRSYCSRSIRSYLREVSYPSTVGA 180  
 DB 124 DLHVNSONIAETRAEDPPWFEGLESFQSGYLRSYCSRSIRSYLREVSATYSWDE 183  
 QY 181 EAQEFPLAVLSCQRLSMOYNGSYFDRGAKGSRCTPGWFSQGPFDMSCLSRHS 240  
 DB 184 AAOEYFLVLSMCKLKSVOYNGSYFDRGAEASRLCTPGWFSQGPFDMSCLSRHS 243  
 QY 241 INPYSNRSRLFTSTWINDHIEKRTIPLVNAIKQDGRVDMEFYLLTSENLK 300  
 DB 244 INPYSNRSRLFTSTWINDHIEKRTIPLVNAIKQDGRVDMEFYLLTSENLK 301  
 QY 301 LVHIVCHKKTTHKLCNCPDSRIYKQPTRLKRPYARRQ 338  
 DB 302 LVHIVCHKKTTHKLCNCPDSRIYKQPTRLKRPYARRQ 339

RESULT 7  
 AA009550  
 ID AA009550 standard; Protein: 96 AA.  
 AC AA009550;  
 DT 06-NOV-2001 (first entry)  
 DE Human polypeptide seq ID NO 23442.  
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;





OY 227 OGFEDMSCLSRHSINPYSNRESRI-----LFTWNLDHI-----IEKKRIIPTL 272  
 Db 390 KMALDIEIATYRKLL---EGESRSISLPFTFSSLNLRNLESLPIVDTHSKRTL---L 443  
 OY 273 VEAIKEDQGREVD 285  
 Db 444 IKVETRDGOVIN 456

RESULT 11  
 ABB57310  
 ID ABB57310 standard; Protein: 466 AA.  
 XX  
 AC ABB57310;

07-MAR-2002 (first entry)

XX Mouse ischaemic condition related protein sequence SEQ ID NO:867.

XX Mouse: Ischaemia; compressive ischaemia; occlusive ischaemia;  
 XX vasoospastic ischaemia; ischaemic condition; ischaemic disease.

XX Mus musculus.

XX W0200188188-A2.

XX 22-NOV-2001.

XX 18-MAY-2001; 2001WO-JP04192.

XX 18-MAY-2000; 2000JP-0145977.

XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

XX WPI: 2002-034733/04.

XX N-PSDB; AB199777.

XX Examining the ischemic condition (e.g. occlusive ischemia) by measuring  
 PT expression levels of particular genes defined in the specification or  
 PT by determining the expression profile of a gene group comprising these  
 PT genes -

XX Claim 2; Page 2178-2180; 2690pp; English.

XX The present invention describes a method for examining ischaemic  
 conditions, comprising measuring the expression levels of particular  
 genes (1) in a test sample or determining the expression profile of a  
 gene group in the sample comprising genes selected from (1). The method  
 is useful for examining the ischaemic condition (e.g. compressive  
 CC ischaemia, occlusive ischaemia or vasoospastic ischaemia) by measuring  
 CC expression levels of particular genes (AB199202 to AB199912, encoding  
 CC the protein sequences in ABB57020 to ABB57374) or by determining the  
 CC expression profile of a gene group comprising these genes. The  
 CC expression levels or expression profiles produced by these genes are  
 CC used as an indicator when screening for ischaemic condition-improving  
 CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914  
 CC represent PCR primers for a mouse ischaemic condition related sequence,  
 CC which are used in the exemplification of the present invention.

XX Sequence 466 AA;

Query Match 5.5%; Score 99; DB 23; Length 466;

Best Local Similarity 20.1%; Pred. No. 0.9; Indels 90; Gaps 14;

Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;  
 OY 27 OEYLKRGCLRPOLPGRSGRLCYEDGTETEDYPPSPVNAELVLTGQMGQVSDIR 86  
 Db 180 EDIMR---LREKIQE-----MLGREAEENTLQSFROVDNASTARLDLKRKVESIQEETI- 231  
 OY 87 RFLSAFHEPOVGLIQAA-----QQLLDEQAPQORILADLHNNSQNTAAETRAEDPPW 141

Db 232 AFLKLIHDEIOELOAOIOEQHVOIDVDSKPDLTALRDVROQY-ESYAARKNIQAEAEW 290  
 OY 142 F-----EGLSRFQSKSGYLRY-----SCESRIK-----S 166  
 Db 291 YKSKFADLSEANRRNDALROAKQESNEYRQVQSITCEVDALKGTNESLERQREMEEN 350  
 OY 167 YLREVSSTYSPTGAEQOEPLRVLGSMQORRSQVNGSFPBGAKGSRLOCTPCQWPC 226  
 Db 351 FALGANVQDTIG-RLODEIQMKKEAARHLREYQ-----DLLNV 389  
 OY 227 OGFEDMSCLSRHSINPYSNRESRI-----LFTWNLDHI-----IEKKRIIPTL 272  
 Db 390 KMALDIEIATYRKLL---EGESRSISLPFTFSSLNLRNLESLPIVDTHSKRTL---L 443  
 OY 273 VEAIKEDQGREVD 285  
 Db 444 IKVETRDGOVIN 456

RESULT 12

AAM54351  
 ID AAM54351 standard; protein: 465 AA.

XX AAM54351;

XX 14-AUG-1998 (first entry)

XX Vimentin.

XX Endometrium; hyperplasia; adenocarcinoma; proliferative phase;

XX 2D gel electrophoresis; detection.

XX Homo sapiens.

XX W09810291-A1.

XX 12-MAR-1998.

XX 05-SEP-1997; 97MO-GB02394.

XX 08-APR-1997; 97GB-0007132.

XX 06-SEP-1996; 96GB-0018600.

XX (CLIN-) CENT CLINICAL & BASIC RES.

XX Byrjalsen I, Rey SJ, Larsen P;

XX WPI: 1998-207057/18.

XX Biochemical markers of human endometrium - useful for, e.g.  
 PT diagnosis of hyperplasia and adenocarcinoma

XX Disclosure; Page 20; 77pp; English.

XX proteins AAM54349-W54364 are examples of proteins produced in the  
 CC endometrium during the hyperplasia, adenocarcinoma or proliferative  
 CC phase of the endometrium. The presence and quantities of these proteins  
 CC can be detected using 2D gel electrophoresis comparison of cell lysates.  
 CC The proteins can be used as biochemical markers to detect the phase of  
 CC the endometrium and can be measured in body fluids, obviating the need  
 CC for endometrial biopsies.

XX Sequence 465 AA;

Query Match 5.4%; Score 97; DB 19; Length 465;

Best Local Similarity 20.8%; Pred. No. 1.4; Indels 90; Gaps 15;

Matches 65; Conservative 46; Mismatches 112; Indels 90; Gaps 15;  
 OY 27 OEYLKRGCLRPOLPGRSGRLCYEDGTETEDYPPSPVNAELVLTGQMGQVSDIR 86  
 Db 179 EDIMR---LREKIQE-----MLGREAEENTLQSFROVDNASTARLDLKRKVESIQEETI- 230  
 OY 87 RFLSAFHEPOVGLIQAA-----QQLLDEQAPQORILADLHNNSQNTAAETRAEDPPW 141

Db 231 AFLKRLHEEIOEQAOIQEOHVOIDVVSFKPDLTALRDVROQY-ESVAARKNLQEAIEW 289  
 142 F-----EGLESRFQSKSGLYR-----SCESRIR-----S 166  
 Db 290 YKSKFADISEANRRNDLRQAKQESTERYRQVOSLCEVDALKTNSLERQKREMEEN 349  
 167 YLREVSSTYPTVGAEOEFLRVLSMCQRLSMQYNGSYPRGAKGGSRLCTPEGWFS 226  
 Db 350 FAVEANAYQDTIG-RLQDEIONMKEMARHLREYO-----DLLNV 388  
 227 QGPPDMSCLSRHSINPYSNRESRL-----FST-----WNLDT-----IEKKRTIPTL 272  
 Db 389 KMALDIEIATYRKLL---EGESRISLPLPWFSSLNLRNRETNLDSLPVDTSHSKRTF---L 442  
 273 VEAIKEDGPREVD 285  
 443 IKVETTRDGOVIN 455

AAB29635 standard; Protein; 466 AA.  
 AAB29635;

21-FEB-2001 (first entry)

Human pollinosis-associated gene 795-encoded protein, SEQ ID NO:26.

Human: pollinosis-associated gene 795; vimentin homologue;  
 IgE; immunoglobulin E; cedar pollen allergy; T-cell; reduced expression;  
 detection; diagnosis; drug screening; allergic disease.

Homo sapiens.

MO200065050-A1.

02-NOV-2000.

26-APR-2000; 2000MO-JP02734.

27-APR-1999; 99JP-0120494.

(GENO-) GENOX R&S INC.

(EISA) EISAI CO LTD.

Nagasu T, Sugita Y, Kashiwabara T, Oshida T, Obayashi M, Gunji S;

Obayashi I, Imai Y, Yoshida N, Ogawa K, Matsui K, Takahashi E;

Yokoi A;

WPI: 2000-687343/67.

N-PSDB: AAC64226.

Page 64-67; Claim 13; 73pp; Japanese.

The invention relates to the human pollinosis-associated gene 795 which  
 exhibits significantly reduced expression in the T-cells of individuals  
 with high cedar pollen-specific IgE (immunoglobulin E) levels. The gene  
 was isolated from T-cells from individuals allergic to cedar pollen using  
 the differential display method. Pollinosis-associated gene 795 has  
 homology with the human vimentin gene. The invention also relates also  
 relates to the protein encoded by pollinosis gene 795; to expression  
 constructs and host cells comprising pollinosis-associated gene 795  
 nucleic acids; pollinosis-associated gene 795 primers and probes;  
 antibodies against the protein encoded by the gene; methods of detection  
 of pollinosis-associated gene 795 nucleic acids; and a method of  
 diagnosis of allergic diseases via the detection of pollinosis-associated  
 gene 795 nucleic acids. The invention additionally encompasses methods of

screening drug candidates for the treatment of allergic disease by  
 measuring the expression of pollinosis-associated gene 795 in pollen  
 antigen-stimulated T-cells in the presence of a test compound relative to  
 a control. Pollinosis-associated gene 795 is useful in the diagnosis of  
 allergic diseases and in the screening of drug candidates for the  
 treatment of such diseases. The present sequence represents a  
 protein encoded by human pollinosis-associated gene 795.

Sequence 466 AA;

Query Match 5.4%; Score 97; DB 21; Length 466;

Best Local Similarity 20.8%; Pred. No. 1.4; Indels 90; Gaps 15;

Matches 65; Conservative 46; Mismatches 112; Indels 90; Gaps 15;

27 QEVLRKGLRFPQLENGSRGLCYEDGETLTDYPPSPDANAEVLTIGAMQGVSDIR 86

180 EDIMR---LRKRLQEE---MLQREEMNTIQSRQVDNMSLARLDERVSELOEETI 231

87 RFLSAFHEPQVGLIOA-----QQLLCDEQAPQKORLIADLHNVSQNIATRAEDPPW 141

232 AFLKRLHEEIOEQAOIQEOHVOIDVVSFKPDLTALRDVROQY-ESVAARKNLQEAIEW 290

142 F-----EGLESRFQSKSGLYR-----SCESRIR-----S 166

291 YKSKFADISEANRRNDLRQAKQESTERYRQVOSLCEVDALKTNSLERQKREMEEN 350

167 YLREVSSTYPTVGAEOEFLRVLSMCQRLSMQYNGSYPRGAKGGSRLCTPEGWFS 226

351 FAVEANAYQDTIG-RLQDEIONMKEMARHLREYO-----DLLNV 389

227 QGPPDMSCLSRHSINPYSNRESRL-----FST-----WNLDT-----IEKKRTIPTL 272

390 KMALDIEIATYRKLL---EGESRISLPLPWFSSLNLRNRETNLDSLPVDTSHSKRTF---L 443

273 VEAIKEDGPREVD 285

444 IKVETTRDGOVIN 456

RESULT 14

AAB66348 standard; peptide; 466 AA.

AAB66348;

05-APR-2001 (first entry)

Human vimentin.

Vimentin; caspase cleavage site; apoptosis; antibody; Basedow's disease;

systemic lupus erythematosus; autoimmune haemolytic anaemia; AIDS; human.

Homo sapiens.

EPI067142-A1.

10-JAN-2001.

07-JUL-2000; 2000EP-0305736.

07-JUL-1999; 99JP-0193235.

(RIKE) RIKEN KK.

WPI: 2001-149349/16.

New antibody reacting with a cleavage product of vimentin but not with  
 the intact vimentin, useful for detecting apoptosis and the quantity of  
 cleavage product of vimentin or as a reagent for immunohistochemical  
 staining



XX 05-FEB-2001; 2001WO-US04098.  
 XX 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI: 2001-476283/51.  
 DR N-PSDB; AAK52970.  
 DR Nucleic acids encoding polypeptides with cytokine-like activities,  
 useful in diagnosis and gene therapy -  
 PS Claim 20; Page 353-354; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM7833-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC  
 XX Sequence 1152 AA;  
 SO  
 Query Match 5.38; Score 94.5; DB 22; Length 1152;  
 Best Local Similarity 21.28; Pred. No. 9.4;  
 Matches 83; Conservative 54; Mismatches 155; Indels 99; Gaps 18;  
 PS 4 KPSVKLRAL-----RSPKFGVAGRSQ-----EYLKRGCLR--FQLEPGRSLCLYE 50  
 DB 489 RPKQELRHLCDFSAEIAIRRGILSKSLHRLLETSMKWCDEGTYLLASQPYDKCQSQ 548  
 CC 51 DGEFLT-EDYFPPVDPNAELVLLTGGAMQGVSDIRRL-----SAFHE 94  
 CC 549 DGAELALQETKELEFETGAENKIOELNAYKEYESILNQDIMEHVAKVFOKASMEVEFHR 608  
 CC 95 POGVL--IQAAQQLCDEQAPQORRLADLLHNQSIATRAEDPPWEGLESRQSK 152  
 DB 609 RQSLKTLARQTRPVQVAPRPAAL-----AKSPDPSGIRGSSNS--SSE 654  
 CC 153 SGYLRYSCESIRISYLRVSSYPTVGAEOEFLRVIGS--MQRRLSMQYNSYDRG 210  
 DB 655 GGAIRRPYRAKSEMSESRQGRSAGE--EESLAILRRHVSLLDTER--AYVE-- 707  
 CC 211 AKGGRSLCTPEGWFGSCGPPDMDCSLRHSINPYSNRESRIETSTWNLDIITE----- 263  
 DB 708 -----ELLCVLEGYAA-----EMDNPLMAHLITGLHNKKVLLG--NMEEIYHNNRIFL 756  
 CC 264 ---KKRIITPLV-----EAIKEDGKREVDWEYFGLLFTSENILKLVHIV 305  
 DB 757 RELENTYDCELVGRCLERMEDEFOIYEKCONKPRSESLMROCSDCPFOE----- 808  
 CC 306 CHKRTTHKNCDPRIYKPPOTRLKRQPVKR 336

DB 809 CQRLDKHLSD-SYLKPVQRTTKYQLLK 838  
 RESULT 17  
 AAM78853  
 ID AAM78853 standard; Protein: 1120 AA.  
 XX  
 AC AAM78853;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human protein SEQ ID NO 1515.  
 XX  
 KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KW tissue growth factor; immunomodulatory; cancer; leukemia;  
 KW nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WO-US04098.  
 XX  
 PR 03-FEB-2000; 2000US-0496914.  
 PR 27-APR-2000; 2000US-0560875.  
 PR 20-JUN-2000; 2000US-0598075.  
 PR 19-JUL-2000; 2000US-0620325.  
 PR 01-SEP-2000; 2000US-0654936.  
 PR 15-SEP-2000; 2000US-0663561.  
 PR 20-OCT-2000; 2000US-0693325.  
 PR 30-NOV-2000; 2000US-0728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;  
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX WPI: 2001-476283/51.  
 DR N-PSDB; AAK51986.  
 DR Nucleic acids encoding polypeptides with cytokine-like activities,  
 useful in diagnosis and gene therapy -  
 PS Claim 20; Page 3811-3813; 6221pp; English.  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM7833-AAM80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activin/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukemia, nervous system disorders, arthritis and  
 CC inflammation.  
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666  
 CC (AAM80020) are omitted as the relevant pages from the sequence listing  
 CC were missing at the time of publication.  
 CC  
 XX Sequence 1120 AA;  
 SO  
 Query Match 5.28; Score 93.5; DB 22; Length 1120;  
 Best Local Similarity 21.28; Pred. No. 11;  
 Matches 83; Conservative 54; Mismatches 155; Indels 99; Gaps 18;  
 PS 4 KPSVKLRAL-----RSPKFGVAGRSQ-----EYLKRGCLR--FQLEPGRSLCLYE 50  
 DB 457 RPKQELRHLCDFSAEIAIRRGILSKSLHRLLETSMKWCDEGTYLLASQPYDKCQSQ 516

```

OY 51 DGETELT-EDYEPSPVDNAELVLTLLGOAMOGYSDIRREPL-----SAFEH 94
Db 517 DGAELALOEIEKEFLETGEMENKIOELNATYKEYESIINODLMEHVKRVEFOKQASMEYFHR 576
OY 95 POGVL--IOAAQOLLCDOAPORORLADLHNVSONIAETRAEDPMPFGLESRSQSK 152
Db 577 RQASLKTKLAAROTRVPQVAPRPREAL-----AKSPCSPCIRGSENS--SSE 622
OY 153 SGYLRSCSRIRSRVSVSYPTVAEAOEFLVIGS--MCORLRSMOYNGSYDRG 210
Db 623 GGLARRGPPYRAKSVSVSSRGSGSAEE--EESLAILRRHVMSELDTER--AYVE-- 675
OY 211 AKGSRICPEEGMEFCOGPPMDSCLSRHSINFSRRESRIILFTWNLPHILE----- 263
676 ----ELLCVLEGIYA-----EMDNPPLMAHLSTGLINKKQVLEG--NMEEIYHFRNRIFL 724
264 ----KKRTIIPILV-----EAIKROGQREVDMETFGLLFTSENLKLVHIV 305
Db 725 RELENTDCPELVGCEFLERNMEDPOLYEKTCQNKPRSESLMQCSDCEPQD----- 776
OY 306 CRRKTHKINCDPSRIYKQTRLEKRPVRK 336
Db 777 CORRIDHKLISD-STYLLKPVQRIYKQVLLK 806

RESULT 18
AA040801
ID AA040801 standard; Protein: 181 AA.
AC AA040801;
XX
XX 22-OCT-2001 (first entry)
DE Human polypeptide SEQ ID NO 5732.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
XX
XX 03-AUG-2000; 2000US-0653450.
XX
XX 14-SEP-2000; 2000US-0662191.
XX
XX 19-OCT-2000; 2000US-0693036.
XX
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSB-) HYSBQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D,
XX Mang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J,
XX Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-44253/47.
XX
XX N-PSDB; AAI59957.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5732; 10078pp; English.

```

```

XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AA038642-AA042213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: immune system suppression,
XX activation/inhibition of activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 181 AA:
XX
XX Query Match 5.2%; Score 93; DB 22; Length 181;
XX Best Local Similarity 24.5%; Pred. No. 0.86;
XX Matches 50; Conservative 22; Mismatches 56; Indels 76; Gaps 11;
XX
XX OY 24 RSCOEVLK-----GCLRFQLPERSRLCIYEDGETELT 56
Db 4 RSCQSPRRKSRRAHVTYLVCGFTSFSFLPYLCGLRF--PE--RTCSQLQADWA 58
OY 57 EDYFSP--VPDNAELVLTLLGOAMOGYSDIRRELSFHPQVGIIOAQQOLLCDOAPQ 114
Db 59 PDFGSPSPV-----SWGATGARKKFLAFN--INLGTREQ----- 94
OY 115 RORLADLHNVSONIAETRAEDPMPFGLESRSQSKSGYLRSCSRIRSYLR--EVS 172
Db 95 -----AHLALNLRREGGRKQD--GLTKVYGIGYIDENKLNQAVSTNLLDEPVT 143
OY 173 S-----YPTVGAEOEFLVIGS 192
Db 144 ALHTVYETIC-REAOELSLPYVGS 166

RESULT 19
AA003204
ID AA003204 standard; Protein: 221 AA.
AC AA003204;
XX
XX 12-SEP-2001 (first entry)
DE Human CIDE protein consensus sequence.
XX
XX HLXAL47; CIDE; Cell death inducing DFF45-like effector;
XX apoptosis; programmed cell death; cancer; haematological disorder;
XX bone marrow failure; myelodysplastic syndrome; aplastic anaemia;
XX neutropenia; leukemia; cardiovascular disease; ischaemia;
XX reperfusion; liver disease; inflammation; neurodegenerative disease.
XX
XX Synthetic.
XX
XX OS Homo sapiens.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
XX FT Misc-difference 70 /label= OTHER
XX FT /note= "Other- unknown"
XX FT Misc-difference 124 /label= OTHER
XX FT /note= "Other- unknown"
XX FT Misc-difference 173 /label= OTHER
XX FT /note= "Other- unknown"
XX FT Misc-difference 185 /label= OTHER

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OK protein - nucleic search, using frame\_plus\_p2n model

Run on: May 26, 2003, 15:21:43 ; Search time 3125 Seconds  
(without alignments)  
3147.759 Million cell updates/sec

Title: US-09-748-451-2  
Perfect score: 1789  
Sequence: 1 MDAQKSVKALALSPKRF.....SRIYPRIRKRPVRRKQ 338

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ p2n.model -DEV=xlh  
-O=/gen2.1/USPTO.spool/US09748451/runat\_21052003\_153829\_16279/app\_query.fasta\_1.519  
-DB=genembl -OPMT=fastap -SUPER=irge -MINMATCH=0.1 -LOOPEXT=0  
-INITS=bits -START=1 -END=1 -MATRIX=biolum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=opt -THR MAX=100 -THR MIN=0 -ALIGN=20 -MODE=LOCAL  
-CURRENT=1 -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09748451 -GCN 1.1 2496 -runat\_21052003\_153829\_16279 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEROVER -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOC  
-DEV\_TIMBOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl.\*  
1: gb\_ba.\*  
2: gb\_hgt.\*  
3: gb\_in.\*  
4: gb\_om.\*  
5: gb\_ov.\*  
6: gb\_pat.\*  
7: gb\_ph.\*  
8: gb\_pl.\*  
9: gb\_pr.\*  
10: gb\_ro.\*  
11: gb\_sts.\*  
12: gb\_sy.\*  
13: gb\_un.\*  
14: gb\_vi.\*  
15: em\_ba.\*  
16: em\_fun.\*  
17: em\_hum.\*  
18: em\_in.\*  
19: em\_mu.\*  
20: em\_om.\*  
21: em\_or.\*  
22: em\_ov.\*  
23: em\_pat.\*  
24: em\_ph.\*  
25: em\_pl.\*  
26: em\_ro.\*  
27: em\_sts.\*  
28: em\_un.\*

pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1789	100.0	1017	9	AB013918 Homo sapi
2	1789	100.0	2839	6	AR122560 Sequence
3	1789	100.0	2839	9	AF064619 Homo sapi
4	1786	99.8	1017	9	AF039210 Homo sapi
5	1759.5	98.4	2926	9	AB028813 Homo sapi
6	1758.5	98.3	2962	9	AB028811 Homo sapi
7	1757	98.2	1028	9	AF409060 Homo sapi
8	1754.5	98.1	1043	9	AF409061 Homo sapi
9	1743	97.4	1071	9	AF409062 Homo sapi
10	1730	96.7	3008	6	AB028912 Homo sapi
11	1416	79.2	1038	6	E36548
12	1416	79.2	1038	6	E36551
13	1416	79.2	1038	10	AB009377 Mus muscu
14	1392	77.8	1316	10	AL353671 Homo sapi
15	1251	69.9	15585	2	AL353671 Homo sapi
16	1065	59.5	142970	2	AL353671 Homo sapi
17	1043	58.3	1499	5	AF406761 Gallus ga
18	941	52.6	1159	5	AF286179 Dantio rer
19	410	22.9	160716	2	AL691523 Homo sapi
20	355.5	19.9	1833	9	AK098413 Homo sapi
21	328.5	18.4	4339	5	AF426316 Dantio rer
22	321	17.9	2118	3	AF149797 Drosophi
23	320	17.9	1401	3	AB036773 Drosophi
24	320	17.9	211505	2	AL806525 Mus muscu
25	295	16.2	167159	3	AC009340 Drosophi
26	295	16.2	261432	3	AB003639 Drosophi
27	294	16.4	18766	2	AC012906 Drosophi
28	124.5	7.0	1851	9	HSY12974 Homo sapi
29	111.5	6.2	1600	9	BC030573 Homo sapi
30	109.5	6.1	59280	9	AL603906 Homo sapi
31	108.5	6.1	6943	10	MM005823 Homo sapi
32	108	6.0	1438	8	PSA12156 Homo sapi
33	106.5	6.0	1165	10	AF136601 Homo sapi
34	106.5	6.0	11645	1	AE002394 Homo sapi
35	106.5	6.0	110000	2	LMFICHR36.22
36	106.5	6.0	172148	3	LMFICHR36.22
37	106.5	6.0	195767	1	NMA72491 Homo sapi
38	106.5	6.0	349980	6	AX044029 Sequence
39	105.5	5.9	1714	8	AF032688 Oryza sat
40	105.5	5.9	184765	2	AC129718 Oryza sat
41	104.5	5.8	161223	2	AC124011 Mus muscu
42	104	5.8	7117	4	P1GAP0B07 Mus muscu
43	104	5.8	7117	4	P1GAP0B02 Mus muscu
44	103	5.8	963	9	HSCIC58BC Mus muscu
45	102.5	5.7	1227	9	H0V1M3 Human vimen

RESULT 1

## ALIGNMENTS

*Deleted*  
PRI 11-AUG-1998

AB013918  
LOCUS AB013918 1017 bp mRNA linear  
DEFINITION Homo sapiens mRNA for CAD, complete cds.  
ACCESSION AB013918  
VERSION AB013918.1 GI:3410908  
KEYWORDS CAD.  
SOURCE Homo sapiens lymphoma cell\_line:KT-3 CDNA to mRNA.  
ORGANISM Homo sapiens  
Mammalia; Euteria; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Euteria; Primates; Carnivora; Hominoidea; Homo.  
1 (sites)  
Mukae, N., Enari, M., Sakahira, H., Fukuda, Y., Inazawa, J., Toh, H. and  
Negata, S.  
Molecular cloning and characterization of human caspase-activated  
DNase  
Proc. Natl. Acad. Sci. U.S.A. 95 (16), 9123-9128 (1998) *Notion 105*  
JOURNAL MEDLINE  
REFERENCE  
AUTHORS  
TITLE  
DIRECT SUBMISSION  
2 (bases 1 to 1017)  
Negata, S.  
Submitted (14-MAY-1998) Shigekazu Negata, Osaka University Medical  
School, B-3, Department of Genetics; 2-2 Yamada-oka, Suita, Osaka  
565-0871, Japan (E-mail: negata@genetic.med.osaka-u.ac.jp,  
Tel:81-6-879-3310, Fax:81-6-879-3319)  
Location/Qualifiers  
1. 1017  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/map="1p36.3"  
/cell\_line="KT-3"  
/cell\_type="Lymphoma"  
1. 1017  
/gene="CAD"  
1. 1017  
/gene="CAD"  
/codon\_start=1  
/product="CAD"  
/protein\_id="BAA32250.1"  
/db\_xref="GI:3410908"  
/translation="MLQPKSVKLRALSPKFGVAGRSCEVLRKGLRQLPERS  
RLCYEDGETELTEYEPSPVNAELVLTLOAGGVYDIBRFISARHEPVOGILYA  
AOOLCDAPORORRLADLHNSONIAATRAEDPWPFGLESRSRPSKSGYLRYSC  
ESRIRSYREVSISPTYGAEAGOEFLVLSKCRKLSMGTNSYIFRKGSGRLC  
TPEGWFCQGPDMDCISLRHSINPYSRERILFSTWNLHIIKKRTIIPITVEAI  
KEQDGRVDMWYFVGLFETSNLKLIVHCHKTKTHKLNCDPSRIYKPTRLKROPY  
RKRO"

BASE COUNT 236 a 293 c 303 g 185 t  
ORIGIN  
Alignment Scores:  
1.9e-159 Length: 1017  
1789.00 Matches: 338  
100.00% Conservative: 0  
100.00% Mismatches: 0  
100.00% Indels: 0  
Gaps: 0  
US-09-748-451-2 (1-338) x AB013918 (1-1017)

QY 1 MetLeuGlnYSProlYSerValLYsLeuAlaLeuArgSerProArgLYsPheGly 20  
DB 1 ATGTCCCAAGAGCCCAAGAGCGTGAAGCTGGGGCCCGCCAGCCAGGAAGTTCCGC 60  
QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLYsGlyCysLeuArgPheGlnLeuPro 40  
DB 61 GTGGCTGGCCGAGAGTGGCAGAGGTGCTGGCCAAAGGCTGTCTCCGCTTCCAGCTCCCT 120  
QY 41 GlnArgGlySerArgLeuCysLeuTYrGlnuAspGlyThrGlnuLeuThrGlnuAspTYrPhe 60  
DB 121 GAGCGCGGTCCCGCTGTGCTGACAGAGATGACGAGGAGCTGACGAGGAAGATATACATTC 180  
QY 61 ProSerValProAspAlaGlnuLeuValLeuLeuThrLeuGlnuAlaThrGlnuGly 80

DB 181 CCCAGCTGTCCCGACAGCCGAGCTGTGCTGTCCACCTTGGGCGAGGCTGGAGGCGC 240  
QY 81 TYrValSerAspAlaArgArpPheLeuSerAlaPheHisGlnuProGlnuValGlyLeuIle 100  
DB 241 TATGTAGCGACATAGGGGCTCTCCAGTGTATTCACAGGCCACAGAGTGGGCTCATC 300  
QY 101 GlnAlaAlaGlnGlnLeuLeuCysAspGlnuAlaProGlnuArgGlnuArgLeuLeuAla 120  
DB 301 CAGCGCCGCCACAGACCTGTGTGTATGAGCAGGCGCCACAGAGCAGAGCGTGTGCT 360  
QY 121 AspLeuLeuHisAsnValSerGlnuAsnIleAlaGlnuThrArgAlaGlnuAspProPro 140  
DB 361 GACCCCTCGACACAGCTGACGAGCAATCGGGCGAGCCGCGGCTGAGAGCCCGCG 420  
QY 141 TRPheGlnuGlyLeuGlnuSerArgPheGlnuSerLYsSerGlyTYrLeuArgTYrSerCys 160  
DB 421 TGGTTGAAGGCTTGGAGTCCCGATTTCAGACCAAGCTTGCTATCTAGATACAGCTGT 480  
QY 161 GluSerArgIleArgSerTYrLeuArgGlnuValSerSerTYrProSerThrValGlyAla 180  
DB 481 GAGAGCCGAGATCCGAGATTACCTGAGGAGGAGTGTACCTTCCACAGGTGGTGTGCG 540  
QY 181 GlnAlaGlnuGlnuPheLeuArgValLeuGlySerMetCysGlnuArgLeuArgSerMet 200  
DB 541 GAGGCTCAGAGGAATTCCTGGGCTCCCTCGCTCCAGTGTCCAGAGGCTCGGTCATG 600  
QY 201 GlnTYrAsnGlySerTYrPheAspArgGlnuAlaLYsGlyLYsArgLeuCysThrPro 220  
DB 601 CAGTACCAATGGCAGCTTACTTGCACAGAGAGCCAGAGGCGGAGCGCTCTGCACACG 660  
QY 221 GlnuGlyTYrPheSerCysGlnuGlyProPheAspMetAspSerCysLeuSerArgHisSer 240  
DB 661 GAAGGCTGTCTCTCTGCGAGGCTCCCTTGTGACATGAGCAGAGTGTATCAAGACATCC 720  
QY 241 IleAsnProTYrSerAsnArgGlnuSerArgIleLeuPheSerThrTYrAsnLeuAspHis 260  
DB 721 ATCAACCCCTCAGAGAACAGGAGAGGAGGATCTCTTACAGACCTGGAACCTGTATC 780  
QY 261 IleIleGlnuLYsLYsArgThrIleLeuProThrLeuValGlnuAlaIleLYsGlnuAsp 280  
DB 781 ATAAATGAAAAGAAACGACCACTATCTCTACACGTGGGAAAGCAATTAAGAACACAT 840  
QY 281 GlnArgGlnuValAspTYrGlnuTYrPheTYrGlyLeuLeuPheThrSerGlnuAsnLeuLYs 300  
DB 841 GGAAGAGAGAGTGGAGTGTATTTATGGCTGTCTTATCTCAGAGAACCTATAA 900  
QY 301 LeuValHisIleValCysHisLYsTYrThrHisLYsLeuAsnCysAspProSerArg 320  
DB 901 CTAGTGCAATGTCTGCTATTAAGAAACCAACCAAGCTCACTGATGACCAAGCAGA 960  
QY 321 IleTYrLYsProGlnuThrArgLeuLYsArgLYsGlnuProValArgLYsArgGln 338  
DB 961 ATCTCAAAACCCACAGCAAGGTGAAGCGGAAGCAGCTGTGGGAAACGCCAG 1014  
RESULT 2  
AR122560 2839 bp DNA linear PAT 16-MAY-2001  
LOCUS AR122560  
DEFINITION Sequence 1 from patent US 6165737.  
ACCESSION AR122560  
VERSION AR122560.1 GI:14106877  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2839)  
AUTHORS Wang, X. and Liu, X.  
TITLE DNA fragmentation factor involved in apoptosis  
JOURNAL Patent: US 6165737-A 1 26-DEC-2000;  
FEATURES Location/Qualifiers  
1. 2839  
/organism="unknown"  
BASE COUNT 644 a 701 c 753 g 739 t

## ORIGIN

## Alignment Scores:

Pred. No.: 7,09e-159 Length: 2839  
 Score: 1789.00 Matches: 338  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 6 Gaps: 0

US-09-748-451-2 (1-338) x AR122560 (1-2839)

QY 1 MetleuGlnlyProlySerVallylsleuAlaIleuArgSerProArglysphegly 20  
 132 ATGCTCCAGAACGCCAAGACGTCGAGGCGCCCTGCGCACCGGAGAGTTCGGC 191  
 21 ValAlaGlyArgSerCysGlnGluValleuArglyGlyCysleuArgpheGlnleuPro 40  
 192 GTGGCTGGCCGAGACTCCAGAGAGTGGCTGGCCAGAGGCTCTCCGCTCCAGACTCCCT 251  
 QY 41 GluArgGlySerArgleuCyseuArglyGluValpGlyThrGluIleuThrGluAspTyrPhe 60  
 252 GAGCCCGGCTTCGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 311  
 QY 61 ProSerValProAspAsnAlaGluIleuValleuIleuArglyGlnAlaArgpGlnly 80  
 312 CCCAGCTGTCGCCGACCAAGCGCGAGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 371  
 QY 81 TyrValSerAspIleuArgArgpheleuSerAlaPheHisGluProGlnValGlyleuIle 100  
 372 TATGTGAGCGACATCAGCGCTTCCTCAGTGCATTCACAGGCCACAGAGTGGGCTCATC 431  
 QY 101 GlnAlaIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 120  
 432 CAGGCCCGCCAGCGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 491  
 QY 121 AspIleuIleuHisAsnValSerGlnAsnIleAlaIleuIleuIleuIleuIleuIleuIleu 140  
 492 GACCTCCTGCACACGTCACCGCAACATCGGGCGGAGACCGCGGCTGAGGACCGCCGC 551  
 QY 141 TrpPheGlnGlyIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 160  
 552 TGGTTTGAAGCTTGGAGTCCCATTCACAGCCAGTCTGGCTGATCTGATCTGATCTGATCTG 611  
 QY 161 GluSerArgIleuArgSerTyrIleuArgGluValSerSerTyrProSerThrValGlyAla 180  
 612 GAGACCGCGATCGGAGTCTGAGGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGGAGTCTGAGG 671  
 QY 181 GluAlaGlnIleuGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 200  
 672 GAGGCTCAGAGGAAATCTCGGGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 731  
 QY 201 GlnTyrAsnGlySerTyrPheAspArgGlyAlaIleuGlySerArgleuCyseuThrPro 220  
 732 CAGTCAATGCGACGCTCTGACAGAGGAGCCAGAGGCGGCGGCTCTGACACCG 791  
 QY 221 GluGlyTrpPheSerCysGlnGlyProPheAspMetAspSerCysleuSerArgHisSer 240  
 792 GAAGCTGGTCTCTGCGCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851  
 QY 241 IleAsnProTyrSerAspArgGluSerArgIleuPheSerThrTyrAsnIleuAspHis 260  
 852 ATCAACCCCTACAGTAAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 911  
 QY 261 IleIleGlnIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleuIleu 280  
 912 ATATATGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 971  
 QY 281 GlyArgGluValAspTyrGlyThrPheArgGlyIleuIleuPheThrSerGluAsnIleuLys 300  
 972 GGAAGAGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGGAGTGG 1031  
 QY 301 LeuValHisIleuValCysHisLysLysThrThrHisLysLysAsnCysAspProSerArg 320

DB 1032 CTATGTCACATGTTCTCTCCATTAAGAAACACCCACAGCTCACTGTGACCGAGACAGA 1091  
 QY 321 IleTyrLysProGlnThrArgleuIleuArglyGlnIleuProValArglyArgGln 338  
 DB 1092 ATCTACAAACCCCGACAGAGGTTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1145  
 RESULT 3  
 AF064019  
 LOCUS  
 DEFINITION  
 Homo sapiens DNA fragmentation factor 40 kDa subunit (PFF40) mRNA, complete cds.  
 ACCESSION  
 AF064019  
 VERSION  
 AF064019.1  
 KEYWORDS  
 SOURCE  
 ORGANISM  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 2839)  
 Liu, X., Li, P., Widiak, P., Zou, H., Luo, X., Garrard, W. T., and Wang, X. The 40-kDa subunit of DNA fragmentation factor induces DNA fragmentation and chromatin condensation during apoptosis Proc. Natl. Acad. Sci. U.S.A. 95 (15), 8461-8466 (1998)  
 MEDLINE  
 PUBMED  
 86357957  
 9671700  
 2 (bases 1 to 2839)  
 Liu, X. and Wang, X.  
 Direct Submission  
 Submitted (07-MAY-1998) Biochemistry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75235, USA  
 FEATURES  
 source  
 1..2839  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 1..2839  
 /gene="PFF40"  
 132..1148  
 /function="nuclease; induces DNA fragmentation and chromatin condensation during apoptosis"  
 /codon\_start=1  
 /product="DNA fragmentation factor 40 kDa subunit"  
 /protein\_id="AAC39920.1"  
 /db\_xref="GI:3347857"  
 /translation="MLOKPKSVTLALASPRKGVAGSGCOEVLKRCGLRPOLEPGS RCIYEDGETREYDPSYDPAEYLITLGOAKGYSDIREDLSAPRPOGLTGA AQLICDQARORRIADLIHNSONTIAPRADPWPREDISPROSGSLRSC EKRINSYAREVSSIPSYGAEAEDEERIVSGRCRISNMQYNSYDRKAGSLC IPEKFSQGFDMDSLSRHSINPISNRESILSTWNIDHIEKRTITPILYVAL KQDEREDVMEFYGLFTSENILVHIVCKRTIKLNCDPSHIVAPQTRLRKQPV RKRO"  
 BASE COUNT 644 a 701 c 755 g 739 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7,09e-159 Length: 2839  
 Score: 1789.00 Matches: 338  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-748-451-2 (1-338) x AF064019 (1-2839)  
 QY 1 MetleuGlnlyProlySerVallylsleuAlaIleuArgSerProArglysphegly 20  
 132 ATGCTCCAGAACGCCAAGACGTCGAGGCGCCCTGCGCACCGGAGAGTTCGGC 191  
 21 ValAlaGlyArgSerCysGlnGluValleuArglyGlyCysleuArgpheGlnleuPro 40  
 192 GTGGCTGGCCGAGACTCCAGAGAGTGGCTGGCCAGAGGCTCTCCGCTCCAGACTCCCT 251

Not in IIDS



	gene	1..2926
	/gene="DfF40 delta"	
CDS:	132..443	
	/gene="DfF40 delta"	
	/codon_start=1	
	/product="DNA fragmentation factor 40 kDa subunit delta"	
	/protein_18="ABP0449.1"	
	/db_xref="GI:13516478"	
	/translation="MLDRPKSVYLRLRSPRFKGVAGNSCOEYLRKGLRPOLPERGS RICTLEDTELETFEPFSVPDNELVLTLLGQAWQMDFCHVSODSLTLLSSCPALV	
BASE COUNT	658 a 726 c 776 g 766 t	
ORIGIN		
Alignment Scores:		
Pred. No.:	4,46e-156	Length: 2926
Score:	1759.50	Matches: 337
Percent Similarity:	92.10%	Conservative: 1
Best Local Similarity:	91.83%	Mismatches: 0
Query Match:	98.35%	Gaps: 29
DB:	9	Indels: 1
US-09-748-451-2 (1-338) x AB028913 (1-2926)		
Oy	1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheLy	20
Dd		
Dd	132 ATGCTCCAAACCAAGCAGGTAAAGTGCGGGCCCTGCCAGCCCGAAGAATTCGC	191
Oy	21 ValAlaGlyArgSerCysGlnGluValLeuArgGlyLysLeuArgPheIleuPro	40
Dd		
Dd	192 GTGGCGGCGGAGCTGGCCAGAGCTCTGCCAAGGCTGTCTCCGTTCCAGCTCCCT	251
Oy	41 GluArgGlySerArgLeuLysCysLeuTyrgluLysPglyThrgluLeuthrgluAspTrphe	60
Dd		
Dd	252 GAGCGGAGTGGCTGGCTGTGTACAGAGATGGACGGAAGCTGACGGAATTACTTC	311
Oy	61 ProSerValProAspAsnAlaGluLeuValLeuLeuthrgluLysAlaLarPrpGlnly	80
Dd		
Dd	312 CCCAGGTCCCGAACACGCGCACTGATGTCTCATCTGTGGCCAGCGCTGGACGGC	371
Oy	81 Tyr-----	81
Dd	::::	
Dd	372 TGGTTTTGCATGTTTCCAGATAGTCTCACACTCTGGGCTCAAGCTGTCTCTCC	431
Oy	82 -----ValSerAspIleArgArpHeuSerAla	91
Dd		
Dd	432 TTGGTCTCTTAAGTGTGGGATTACAGANTGATGAGCAGCATCGAGGCTTCTCTAGTCA	491
Oy	92 PheHISgluProGlnValGlyLeuIlEGlAlaAlaGlnIleuLeuLysAspGluIn	111
Dd		
Dd	492 TTTCACAGACCACAGGTGGGCTCATCCAGGCGCCAGAGAGTGTGTATGAGCAG	551
Oy	112 AlaProGlnArgGlnArgLeuLeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAla	131
Dd		
Dd	552 GCCCCACAGAGCCANAGGCTCTGTGCTACCTCTCTCACACATCGACCCAGAACATTCGC	611
Oy	132 AlaGluThrArgAlaGluLysAspProTrpPheGluGlyLeuGluSerArpHeuGlnser	151
Dd		
Dd	612 GCCGAGCCCCGGCTGAGAGACCCCGCGGTGTTGAAGCTTGAAGCTCCGAGTTTACAGC	671
Oy	152 LysSerGlyTYrLeuArgTYrSerCysGluSerArgIleArgSerTYrLeuArgIluVal	171
Dd		
Dd	672 AAGTGGGTACTATGATACAGCTGTGAAGCCGAGATCCGGAGTTACTGTAGGAGAGTG	731
Oy	172 SerSerTYrProSerThrValGlyAlaGluAlaGlnGluLysPheLeuArgValLeuGly	191
Dd		
Dd	732 AGCTCTTACCCCTCCACAGTGGGCGGAGAGCTCAGAGAGAAATTCCTGGCGGCTCTCCGC	791
Oy	192 SerMetCysGlnArgLeuArgSerMetGlyLysArgIlySerTYrPheAspArgIlyAla	211
Dd		
Dd	792 TCCATGTGTCCAAAGAGCTCCGGTCCATCAGTACAAAGGAGAGCTACTCTGCACAGAGACCC	851
Oy	212 LysGlyGlySerArgLeuCYsthProGluGlyTYrPheserCysGlnIlyProphasp	231

Db	852	AAGGGCGGACGCCCTGTGCACACCGGAAGCGTGGTTCTCTCCAGAGGTCCTTTGAC	911
Oy	232	MetLspSerCysLeuSerArgHisSerIleAsnProTyrSerAsnArgIleSerArgIle	251
Oy	912	ATGCAAGACGTGCTTATCAAGACTCCATCAACCCTCACACTAACAGGAGCAGCATC	971
Oy	252	LeuHeserThrTrpAnlnLeuAspHisIleIleGluIlyslYsAsgThrIleIleProThr	271
Oy	972	CTCTTCAGACCTTGAACTTGATGACATATATATAAANAAGAACCCATCATTTCTACA	1031
Oy	272	LeuValGluaIleIleYsgIugInAspGIYargGluValasPTTPGtuyrPheTyrgIy	291
Db	1032	CTGGTGGAGCAATTAGAAGAACAGATGGAAGAGAGTGCGACTGGGATTTTTTAAGGC	1091
Oy	292	LeuLeuPheThrSerGluAsnLeuIylsLeuValHisIleValCySHIsYlsYsthrThr	311
Db	1092	CTGCTTTTAACTCAGAGAACCTTAAACTAGTGCACATTTCTGCCATTAAGAAAAACACC	1151
Oy	312	HisIylsLeuAsnCysAspProSerArgIleThyrIlysProGlnThrArgLeuLyArgIyls	331
Db	1152	CACAGCTCAACTGTGACCCGAGAGAGATTTACAAACCCGACAGCAAGGTGAACGGAG	1211
Db	332	GlnProValArgIylsArgGln 338 	
Db	1212	CAGCTGTGGCGAAGCCGAC 1232 	
RESULT 6	AB028911	2962 bp mRNA linear PRI 03-APR-2001	
LOCUS	AB028911	Homo sapiens DFF40 beta mRNA for DNA fragmentation factor 40 kDa	
DEFINITION	AB028911	subunit beta, complete cds.	
ACCESSION	AB028911.1	GI:13516474	
VERSION	AB028911.1	DNA fragmentation factor 40 kDa subunit beta.	
KEYWORDS		Homo sapiens tissue _lib:Fetal brain CDNA to mRNA.	
SOURCE		Homo sapiens	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
REFERENCE		1 Nakagawara,A., Takahashi,M., Takada,N. and Kawamoto,T. DFF40 beta	
JOURNAL		Published Only in Database (2001)	
AUTHORS		2 (bases 1 to 2962)	
TITLE		Nakagawara,A., Takahashi,M., Takada,N. and Kawamoto,T. Direct Submission	
REFERENCE		Submitted (14-JUN-1999) Akira Nakagawara, Chiba Cancer Center Research Institute, Division of Biochemistry, 666-2 Nitona, Chuo-kai, Chiba, Chiba 260-8717, Japan (E-mail:akiranaka@chiba-cs.pref.chiba.jp, Tel:81-43-264-5431(ex.5201), Fax:81-43-265-4459) Location/Qualifiers	
FEATURES		1..2962	
source		/organism="Homo sapiens"	
gene		/db_xref="taxon:9606"	
CDs		/tissue_1lb="Fetal brain"	
		1..2962	
		/gene="DFF40 beta"	
		132..944	
		/gene="DFF40 beta"	
		/codon_start=1	
		/product="DNA fragmentation factor 40 kDa subunit beta"	
		/protein_id="BAB40447.1"	
		/db_xref="GI:13516475"	
		/translation="MLQPKSVKRLALRSPPKFGVAGSRSCOEVLRKLCARLPDPERSGS RLICLEDGETEDTEYPSPVDMALVELITLGOAMGVSDIRFLSAFHEPOVALIOA AOOLICDEOAGROBORILADILAHNLSOUILLATETRAEDPDMPWGLESRSRGYIRSGC ESRRSTLRKYVSITPSTVGADAQEELFRIVASGMQRKRSMDYNSTIDRCKAGGSRLIC TPKEWFCQCGFDMSCLSRHSIMPYSNRSLIFSTWNLDHDIGVLLCGPG"	
BASE COUNT	668 a	738 c	790 g
ORIGIN			766 t

Pred. No.:	5,63e-156	Length:	2962
Score:	1758.50	Matches:	338
Percent Similarity:	89.18%	Conservative:	0
Best Local Similarity:	89.18%	Mismatches:	41
Query Match:	98.30%	Indels:	1
DB:	9	Gaps:	1

  

US-09-748-451-2 (1-338) x AB028911 (1-2962)	
QY 1 MetLeuGlnIysPProIySserValIlysLeuArgAlaLeuArgSerProAlaGlySPhgIy 20	
Db 132 ATGCTCCAGAAAGCCCAAGAGCGTGAACCTGGGGCCCTGGCCACCCGAGGAAGTTCTGGC 191	
QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgIysGlyCysLeuAcGPhgInLeuPro 40	
Db 192 GTGGCTGGCCGGAAGCTGCCAGAGGTGTCTGGCAGAGGCTGTCTCCGCTTCAAGCTCCCT 251	
QY 41 GluArgIysSerArgLeuCysLeuTrGIuAspGIyThrGluLeuThrGIuAspTrpIyPhe 60	
Db 252 GAGCGCGGTTCCCGGCTGTGGCTGTACAGAGATGGACGAGCGTGAAGGAAGTAACTTCTC 311	
QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGly 80	
Db 312 CCCAGTGTCCCGCAACAGCCGAGCTGTGTGCTCACTTGGGCGCAGGCTGGCAGAGCC 371	
QY 81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValAlaGlyLeuIle 100	
Db 372 TATGTGACGACATCAGAGCGCTTCTCTCATGTCAATTTCACGAGCCACAGTGGGCTATTC 431	
QY 101 GlnAlaAlaGlnGlnLeuLeuCysAspGluGlnAlaProGlnArgGlnArgLeuAla 120	
Db 432 CAGCGCCCGCCAGCAGCTGTGTGTGATGACAGAGCCCGCACAGAGCAGAGGCTCTGGCT 491	
QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaIleAlaGluThrArgAlaGluAspPro 140	
Db 492 GACCTTCCTGCACCAAGCTCAGCCAGAACATCGCGCCGAGACCCGGGCTGAGAGCCCGCG 551	
QY 141 TrpPheGlnGlyLeuGlnSerArgPheGlnSerIysSerGIyTyrLeuArgTyrSerCys 160	
Db 552 TGTGTTGAAGCTTGGAGTCCCGATTTCAGAGCAAGTCTGGCTATCTGAGATTAAGCTGT 611	
QY 161 GluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAla 180	
Db 612 GAGAGCCGGAATCCGAGATTACTGAGGAGGTGAGTCTACCCCTCCACAGTGGGCTCG 671	
QY 181 GluAlaGlnGlnGluPheLeuArgValLeuGlnSerMetCysGlnArgLeuArgSerMet 200	
Db 672 GAGGCTCAGGAGGAATTCCTGGGGCTCTGGCTCCACATGTCCAGAGGCTCCGGTCCATG 731	
QY 201 GlnIyrAsnGlySerTyrPheAspArgGlyAlaIylsGlyIysSerArgLeuCysThrPro 220	
Db 732 CAGTACATAGGCACACTACTTGCACAGAGAGCAAGAGGCGGCACCGCTGTGCACACGG 791	
QY 221 GluGIyTrpPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240	
Db 792 GAAAGCGTGTCTCTCTGCACAGGTCCTCTTGACATGTGACACTCTTATCAAGACACTCC 851	
QY 241 IleAsnProIyrSerAsnArgGlnSerArgIleLeuPheSerThrTrpPheAlaAspHis 260	
Db 852 ATCAACCCCTACAGTAAACAGGAGAGCAAGATCTCTTACACACTGGAACTGGATCAC 911	
QY 261 Ile----- 261	
Db 912 ATAGATGAGAGTCTGTCTGTGGCCACAGGCTAAAGTGAGGAGATCATGTGCTCACT 971	
QY 261 ----- 261	
Db 972 GAAAGCTGAACCTCTGGGCTTCAGCCGATCTCTGCTCAAGCTCTGAGTGGCTGGGAC 1031	
QY 262 -----1legIuIyIysArgTrhIleIleProthrIleuValGluAlaIleIyGlnGln 279	
Db 1032 TACAGATATGAAAGAAAGCAGCATATCTTCACTCGTGGGAAAGCATTTAAAGAACAA 1091	

QY 280 ASPLVARGLVAlValAspTrpGluTyrPheTyrGlyLeuLeuPheTherSerGluLysAsnLeu 299  
 DB 1092 GATGAGAGAGAGAGAGAGAGAGATTTTATAGGCGCTGCTTTTACCTCAGAGAACTTA 1151  
 QY 300 LysLeuValHisIleValCysHisIleValThrThrHisIleValAsnCysProSer 319  
 DB 1152 AAATCTGTCACATTTGTGTGTCATTAAGAAACCAACCAAGTCAACGTGACCCGAG 1211  
 QY 320 ArgIleTyrLysProGlnThrArgLeuLysArgGlnProValArgLysArgGln 338  
 DB 1212 AGAATCTCAAAACCCCAAGAGTGAAGCGAGAGCGCTGTGCGAAGCCAG 1268  
 RESULT 7  
 AF409060 1028 bp mRNA linear **Ex 11-SEP-2001**  
 LOCUS Homo sapiens DNA fragmentation factor B truncated form I (DFFB)  
 DEFINITION AF409060  
 VERSION AF409060.1 GI:15553728  
 KEYWORDS  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 1028)  
 AUTHORS Bayasas,J.R., Yuste,V.J., Perera,R. and Comella,J.X.  
 TITLE Characterization of splice variants of human caspase-activated  
 DNase that show CIDF-N structure and function  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1028)  
 AUTHORS Comella,J.X., Bayasas,J.R., Yuste,V.J. and Perera,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-AUG-2001) Ciencies Mediques Basiques, Universitat de  
 Lleida, Av. Rovira Roure 44, Lleida 25198, Spain  
 FEATURES  
 SOURCE Location/Qualifiers  
 1..1028  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="IMR-5"  
 /cell\_type="neuroblastoma"  
 1..1028  
 /gene="DFFB"  
 5..337  
 /gene="DFFB"  
 /note="CIDF-N-only protein; alternatively spliced"  
 /codon\_start=1  
 /product="DNA fragmentation factor B truncated form I"  
 /protein\_id="AA02005.1"  
 /db\_xref="GI:15553728"  
 /translation="MIDPKKSVKRLARSPKRGVAGSCQEVLRKGLRQLPERSG  
 RLYLNDGHELEDFRFSVPDMALVLTITGCMQCGCERHQLPQCISRTGCAH  
 PGRPAAY"  
 BASE COUNT 238 a 295 c 308 g 187 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,01e-156 Length: 1028  
 Score: 1757.00 Matches: 337  
 Percent Similarity: 98.83% Conserved: 0  
 Best Local Similarity: 98.83% Mismatches: 1  
 Query Match: 98.21% Indels: 3  
 DB: 9 Gaps: 1  
 US-09-748-451-2 (1-338) x AF409060 (1-1028)  
 QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20  
 DB 5 ATGCTCCAGAACGCCAAGAGCGTGAAGCTGGCGGCCCTGGCGAGCCGAGAAAGTTCCGC 64  
 QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyLysLeuArgPheGlnLeuPro 40  
 DB 65 GTGGCTGGCGGAGAGTGCAGAGAGAGTGTGCGCAAGGAGCTGTCCGCTTCCAGAGCTCCT 124

QY 41 GluArgLysSerArgLeuCysLeuArgLysAspGlyThrGluLeuThrGluLysAspTrpPhe 60  
 DB 125 GAGCGCGGCTTCCGGGCTGTGCTGCTTACAGAGATGGACGAGGCTGACGAGAAATTAATTC 184  
 QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaTrpGlnGly 80  
 DB 185 CCGAGTGTCCCAACACCCGACGCTGTGTCTCACTTGGCCAGGCTGTGGCGAGGCG 244  
 QY 81 -Tyr-----ValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnVal 98  
 DB 245 TGTGGCAGATGTAGGACATGACGCGCTTCTCTAGTGCATTTCAGACGCCACAGGTGGG 304  
 QY 98 YLeuIleGlnAlaIleGlnGlnLeuLeuLeuCysAspGluGlnAlaProGlnArgGln 118  
 DB 305 GCTCATCCAGCCGCCGCCACAGCTCTGTGTGATGACAGGCCGCCACAGAGCAGAGCT 364  
 QY 118 ULeuAlaAspLeuLeuHisValSerGlnAsnIleAlaAlaGluThrArgAlaGlu 138  
 DB 365 GCTGCTGACCTCTCTCAACAGTCAGACAGAACATCGCGGCGAGACCCGGGCTGAGGA 424  
 QY 138 PProTrpTrpPheGluGlyLeuGluSerArgPheGlnSerLysSerGlyTyrLeuArgTyr 158  
 DB 425 CCGCGCGTGTGGAAGGCTGTGAGTCCGATTCAGAGCAAGTCTGGGCTATCTGAGATA 484  
 QY 158 rSerCysGluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThy 178  
 DB 485 CAGCTGTGAGACCGGATCCGAGATTACTGAGGAGGTGAGCTTCACTCCACGCT 544  
 QY 178 IGLYAlaGluAlaGlnGlnGluPheLeuArgValLeuGlySerMetCysGlnArgLeu 198  
 DB 545 GGGTGGAGAGCTCAGAGAGAAATCTCTGGGGTCTCTGCTCCATGTGCCAGAGGCTCG 604  
 QY 198 gSerMetGlnTyrAsnGlySerTyrPheAspArgGlyAlaLysGlyLysArgLeu 218  
 DB 605 GTCCATCATCAACAAAGCGAGCTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 664  
 QY 218 sThrProGluGlyTyrPheSerCysGlnGlyProPheAspMetLysSerCysLeuSer 238  
 DB 665 CACACCGGAGAGGCTGTCTCTCCAGAGGCTCCCTTGAATGACAGTGCAGCTTATCAAG 724  
 QY 238 gHisSerIleAsnProTyrSerAsnArgGluSerArgIleLeuPheSerThrTrpAsn 258  
 DB 725 ACATCTCATCAACCCCTACAGTAAGGAGAGAGAGAGATCTCTTCAAGACCTGAGACT 784  
 QY 258 uAspHisIleIleGlnLysLysArgThrIleIleProThrLeuValGlnAlaIleLys 278  
 DB 785 GGATCACAATATATGAAGAAAGCAACCATCTATCTCACTGCTGTAACCAATTAAAGA 844  
 QY 278 uGlnAspGlyArgGluValAspTrpGluTyrPheTyrGlyLeuLeuPheTherSerGlu 298  
 DB 845 ACAAGATGAGAGAGAGAGAGAGAGATTTTATGAGCTGCTTTTACCTCAAGAGA 904  
 QY 298 nLeuLysLeuValHisIleValCysHisIleLysLysThrThrHisLysLeuAsnCysAsp 318  
 DB 905 CTTAAACTGTGACATTTGTCTGCCATTAAGAAACCAACCAAGCTCAACTGTATACC 964  
 QY 318 oSerArgIleTyrLysProGlnThrArgLeuLysArgGlnProValArgLysArgGly 338  
 DB 965 AAGGAGATCTCAAAACCCCAAGAGAGTGAAGCGAAGAGAGCTGTGCGAAGCCCA 1024  
 QY 338 n 338  
 DB 1025 G 1025  
 RESULT 8  
 AF409061 1043 bp mRNA linear **Ex 11-SEP-2001**  
 LOCUS Homo sapiens DNA fragmentation factor B truncated form II (DFFB)  
 DEFINITION AF409061  
 VERSION AF409061.1 GI:15553730  
 KEYWORDS  
 SOURCE Homo sapiens.

ORGANISM Homo sapiens  
 Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.  
 REFERENCE 1 (bases 1 to 1043)  
 AUTHORS Bayasacas,J.R., Yuste,V.J., Perera,R. and Comella,J.X.  
 TITLE Characterization of splice variants of human caspase-activated Dnase that show CIDE-N structure and function  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1043)  
 AUTHORS Comella,J.X., Bayasacas,J.R., Yuste,V.J. and Perera,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-AUG-2001) Clencies Mediques Basiques, Universitat de Lleida, Av Rovira Route 44, Lleida 25198, Spain  
 FEATURES  
 source  
 1..1043  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="IMR-5"  
 /cell\_type="neuroblastoma"  
 1..1043  
 /gene="DFFB"  
 5..352  
 /gene="DFFB"  
 /note="CIDE-N only protein; alternatively spliced"  
 /codon\_start=1  
 /product="DNA fragmentation factor B truncated form II"  
 /protein\_id="AA02006.1"  
 /db\_xref="GI:15553731"  
 /translation="MLQPKSVKRLRSPKRGVAGSCOEVLRKCLPQLPERS  
 RLCLYDGETLEDYEDPPSVPDNAELVLTFLGAMQCEMGLMRCERHMLPQCISRA  
 TGAHPGRPAAY"  
 BASE COUNT 242 a 296 c 314 g 191 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 3,52e-156 Length: 1043  
 Score: 1754.50 Matches: 337  
 Percent Similarity: 97.40% Conservative: 0  
 Best Local Similarity: 97.40% Mismatches: 1  
 Query Match: 98.07% Indels: 8  
 DB: 9 Gaps: 1  
 US-09-748-451-2 (1-338) x AF409061 (1-1043)  
 QY 1 MetLeuGlnIySeryProIySserValIySleuArGAlaIeuarGserProArGlySphGly 20  
 DB 5 ATGCTCCAGAAAGCCCAAGAGCGTGAAGCTCGGGCCCTGGCAGCCCGAGAGATTGCGG 64  
 QY 21 ValAlaGlyArGserCysGlnIyValIeuarGlySlyCysIeuarGpHeGlnIeuarPro 40  
 DB 65 GTGGCGTGGCCGAGAGCTGCCAGAGAGCTGCTCCGCTCCGCTCCAGAGCTCCCT 124  
 QY 41 GluArGlySerArGleuCysIeuarGlySlyGlnIyThrGlnIeuarThrGlnIySph 60  
 DB 125 GAGCGGGGTCCCGGCTGCTGCTGACGAGATGGCAGGAGTGAAGAGATTACTTC 184  
 QY 61 ProSerValIyProSpsAnaIyGlnIeuarValIeuarThrIeuarGlnIySphGlnIy 80  
 DB 185 CCCAGTGTCCCAACAAGCCGAGAGCTGCTCCCTGAGGAGGCTGGCCAGGAGG 244  
 QY 81 -Tyr-----ValSerAspIleArGArGpHeuSerAlaPheH 93  
 DB 245 TGTGAGTGGCAAGACTTTGGAGATGTGAGCGACATCAGCGCTTCTCAGTGATTTCA 304  
 QY 93 SglIyProGlnIyValIyLeuIleGlnAlaIyGlnIyLeuIleuIySphIyGlnIyAlaP 113  
 DB 305 CGAGCCACAGGTGGGCTCATCCAGGCGCCAGAGCTCTGTGTATGAGCAGAGGCC 364  
 QY 113 oGlnArGlnArGlyLeuAlaIySphIeuarHIsAsnValSerGlnAsnIleAlaIyG 133  
 DB 365 ACAGAGGAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 424  
 QY 133 uThrArGAlaIySphIySphIySphIySphIySphIySphIySphIySphIySphIyS 153

DB 425 GACCCGGCTGAGAGACCCGCTGTTGAAGCTTGAGCTTGAGATCCGATTTAGACCAAGTC 484  
 QY 153 TGLTYLeuArGlySeryCysGlnIySleuArGlySeryIySleuArGlySleuArGlyS 173  
 DB 485 TGGTATCTAGATAGACGCTGTGAGACCGGATCCGAGATTTACTGAGGAGAGGAGCTC 544  
 QY 173 rTyProSerThrValIyAlaIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIyGlnIy 193  
 DB 545 CTACCCCTCCACGCTGGTGGGAGGCTCAGAGAGATTCCTGCGCTCCCTCCGCTCCAT 604  
 QY 193 tCyGlnArIeuarGserIeuarGlnIyArGlnIySeryIySphIySphIySphIySphIy 213  
 DB 605 GTGCGAGAGGCTCGGCTCAGACGATGACATGAGAGCTGCTGCTGCTGCTGCTGCTGCT 664  
 QY 213 yGlySerArGleuCysThrProIyGlnIyTyrPheSerCysGlnIyProPheSphMetas 233  
 DB 665 CGGAGCGGCTCGGCTGACACCGAGAGCTGTTCTCTGCGAGGCTCCCTTTGACATGA 724  
 QY 233 pSerCysLeuSerArGlnIySeryIySphIySphIySphIySphIySphIySphIy 253  
 DB 725 CAGCTGCTTATCAAGACATCCATCAACCCCTACAGTACAGGAGCAGAGATCCCTCT 784  
 QY 253 eSerThrTrpAsnIeuarSphIleIleGlnIySlyArGlyThrIleIleProthIeuar 273  
 DB 785 CAGCAGCTGGAGACCTGATCACAATATGAAAGAAAGCAACATTCCTACACTGCT 844  
 QY 273 IGlualIleIySglIyGlnIySphIyArGlyIyAlaIySphIyTyrPheIyGlyLeu 293  
 DB 845 GGAAGCAATTAAGAAACAAGTGAAGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 904  
 QY 293 uPheThrSerGlnAsnIySleuValIleIleValIySlySlySlySlySlySlySly 313  
 DB 905 TTTTACCTCAGAGAACTTAAACATGACATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 964  
 QY 313 sIeuarCysAspProSerArGlyIyTyrIySphIyProGlnIyThrArGleuIySphIyS 333  
 DB 965 GCTCACTGTGACCCAGCAAGATCTCAACCCAGACAGAGTTGAAGGAGAGAGGCC 1024  
 QY 333 oValArGlySlyGlnIy 338  
 DB 1025 TGTGCGGAAGCCGAG 1040  
 RESULT 9  
 AF409062 1071 bp mRNA linear PRI 11-SEP-2001  
 LOCUS  
 DEFINITION Homo sapiens DNA fragmentation factor B truncated form III (DFFB)  
 ACCESSION AF409062  
 VERSION AF409062.1 GI:15553732  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens.  
 Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi: Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.  
 REFERENCE 1 (bases 1 to 1071)  
 AUTHORS Bayasacas,J.R., Yuste,V.J., Perera,R. and Comella,J.X.  
 TITLE Characterization of splice variants of human caspase-activated Dnase that show CIDE-N structure and function  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 1071)  
 AUTHORS Comella,J.X., Bayasacas,J.R., Yuste,V.J. and Perera,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-AUG-2001) Clencies Mediques Basiques, Universitat de Lleida, Av Rovira Route 44, Lleida 25198, Spain  
 FEATURES  
 source  
 1..1071  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_line="IMR-5"  
 /cell\_type="neuroblastoma"  
 1..1071  
 /gene="DFFB"

## CDS

5. 247  
 /gene="DFRB"  
 /note="CIDE-N-only protein; alternatively spliced"  
 /codon\_start=1  
 /product="DNA fragmentation factor B truncated form III"  
 /protein\_id="FAL02007.1"  
 /db\_xref="GI:15553733"  
 /translation="MLQPKSVTLRLNSPRKGVAGRSCEVLRLKGLRFOLPERGS  
 RCLYEDGTETLEDYFSPVDPNAELVLTLLGQAWOG"

BASE COUNT 242 a 306 c 319 g 204 t  
 ORIGIN

## Alignment Scores:

4.43e-155 Length: 1071  
 1743.00 Matches: 337  
 94.93% Conservative: 0  
 94.93% Mismatches: 1  
 97.43% Indels: 18  
 Gaps: 1

US-09-748-451-2 (1-338) x AF409062 (1-1071)

OY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20  
 DB 5 ATGCTCCAGAAAGCCCAAGAGCGTAAAGCTGGGCCCCCTGCCAGCCAGAAAGTTGGC 64  
 OY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40  
 DB 65 GTGGCTGGCCGAGGTCCAGAGAGTCTCTCGCAAGAGCGTCTCCGCTCCAGCTCCCT 124  
 OY 41 GluArgGlySerArgLeuCysLeuArgLysPheGlnLeuArgLysPheGlnLeuArgLys 60  
 DB 125 GAGCGCGTTCCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 184  
 OY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaLeuArgLys 80  
 DB 185 CCAGAGTTCCCGAAGAGCGAGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 244  
 OY 81 Tyr-----ValSer 83  
 DB 245 TG-ATGAGCTTCTGCTGAGACCGGACCTTTGTTGCTCCATTGGTGGCAGATGTGAGC 303  
 OY 84 AspLeuArgArgPheLeuSerLysPheLeuArgLysPheGlnLeuArgLysPheGlnAla 103  
 DB 304 GACATCAGGCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 363  
 OY 104 GlnGlnLeuLeuCysAspGlnGlnAlaProLysArgLysPheGlnLeuArgLysPheGln 123  
 DB 364 CAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 423  
 OY 124 HisAsnValSerGlnAsnLysAlaAlaGlnThrArgAlaGlnAspProProTrpPheGln 143  
 DB 424 CACAAAGCTCAGCAAACTGCGGCGGAGAGCCGCGGCTGAGAGCCGCGGCTGAGAGCCG 483  
 OY 144 GlyLeuGlnSerArgPheGlnSerLysSerLysLysLysLysLysLysLysLysLysLys 163  
 DB 484 GCGTTGGAGTCCGATTTTCAGAGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 543  
 OY 164 IleArgSerLysLeuArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 183  
 DB 544 ATCCGAGATTACTGAGGAGGTGAGTCTTACCCCTCCAGCGTGGGTGCGGAGGCTGAG 603  
 OY 184 GlnGlnPheLeuArgValLeuGlnLysSerMetCysGlnArgLeuArgSerMetGlnLysAsn 203  
 DB 604 GAGGAATTCCTGCGGGGCTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 663  
 OY 204 GlySerLysPheAspArgLysLysLysLysLysLysLysLysLysLysLysLysLysLys 223  
 DB 664 GCGACACTACTTCAGACAGAGAGCCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 723  
 OY 224 PheSerCysGlnLysProPheAspMetLysPheCysLysLeuArgLysLysLysLysLys 243  
 DB 724 TTTCTGCTCCAGGAGTCTTTGACATGACAGCTGCTTATCAGACACTCCATCAACCC 783

OY 244 TyrSerAsnArgGluSerArgLysLeuPheSerThrTrpAsnLeuAspHisLysLysLys 263  
 DB 784 TACAGTACAGGAG 843  
 OY 264 LysLysArgThrLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 283  
 DB 844 AAGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 903  
 OY 284 ValAspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLysLeuLysHis 303  
 DB 904 GTGAGCTGGAGTATTTTATGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 963  
 OY 304 IleValCysHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 323  
 DB 964 ATTGCTGCGCAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1023  
 OY 324 ProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338  
 DB 1024 CCCAGACAAAGGTTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1068

Too late

RESULT 10  
 AB028912  
 LOCUS  
 DEFINITION Homo sapiens DFF40 gamma mRNA for DNA fragmentation factor 40 kDa subunit gamma, complete cds.  
 ACCESSION AB028912.1 GI:13516476  
 VERSION DFF40 gamma; DNA fragmentation factor 40 kDa subunit gamma.  
 KEYWORDS Homo sapiens tissue; testis; testis cDNA to mRNA.  
 SOURCE Homo sapiens  
 ORGANISM Homo sapiens  
 REFERENCE 1 (sites)  
 AUTHORS Nakagawa, A., Takahashi, M., Takada, N. and Kawamoto, T.  
 TITLE DFF40 gamma  
 JOURNAL Published Only in Database (2001)  
 REFERENCE 2 (bases 1 to 3008)  
 AUTHORS Nakagawa, A., Takahashi, M., Takada, N. and Kawamoto, T.  
 TITLE Direct Submission  
 JOURNAL Submitted (14-JUN-1999) Akira Nakagawa, Chiba Cancer Center Research Institute, Division of Biochemistry, 666-2 Mitona, Chono-ku, Chiba, Chiba 260-8717, Japan (E-mail: akirana@chiba-cc.pref.chiba.jp, Tel:81-43-264-5431(ex.5201), Fax:81-43-265-4459)

## FEATURES

source 1..3008  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /issue\_11b="testis"  
 1..3008  
 /gene="DFF40 gamma"  
 132..482  
 /gene="DFF40 gamma"  
 /codon\_start=1  
 /product="DNA fragmentation factor 40 kDa subunit gamma"  
 /protein\_id="BAB40449.1"  
 /db\_xref="GI:13516476"  
 /translation="MLQPKSVTLRLNSPRKGVAGRSCEVLRLKGLRFOLPERGS  
 RCLYEDGTETLEDYFSPVDPNAELVLTLLGQAWOGSVGVRAKTRFDTSLSFGDCQ  
 ALDNGRCQRLFL"

## BASE COUNT

673 a 739 c 815 g 791 t

## ORIGIN

Alignment Scores:  
 Pred. No.: 2.8e-153 Length: 3008  
 Score: 1730.00 Matches: 337  
 Percent Similarity: 85.32% Conservative: 0  
 Best Local Similarity: 85.32% Mismatches: 1  
 Query Match: 96.70% Indels: 57  
 DB: 9 Gaps: 1

US-09-748-451-2 (1-338) x AB028912 (1-3008)

OY	1	MelleuglnlyrProlysSerVallylsLeuAargAlaLeuAArgSerProArglysPhegyl	20
Db	132	ATGCTCCAGAAACCCCAAGAGCCTGAAGCTGGCGCCCTCCGACGCCGAGAAAGTTCGGC	191
OY	21	ValAlaGlyAArgSerCysglnGluValLeuAArglyGlyCysLeuAArgPheglnLeuPro	40
Db	192	GGGGCTGGCGCGAGAGCTGCCAGAGAGGGTGGCCCAAGGGGTGTCTCCGCTTCCAGCTCCCT	251
OY	41	GluAArglySerAArgLeucysLeuTyrglyuAspGlyThrGlnLeuThrGlnAspTyrrpe	60
Db	232	GAGCGCGGTCCCGGCTGTGCTGTACAGAGATGGCACCGAGCTGACGGAAAGTATCTTC	311
OY	61	ProSerValProAspAsnAlaGlnLeuValLeuLeuThrLeuGlyGlnAlaArgglnGly	80
Db	312	CCCAAGTGTCCCGACACAGCCGAGGCGTGTCTACACTTGAGGCCAGAGCTGGCAGGGC	371
OY	81	-Tyr-----	81
Db	372	TCTGTGGTGTACAGGCAAGACAAAGACCCGGGACCTCAAGTCTGAGTCTGGTAT	431
	81	-----	81
	432	TGCCAGGCCCTGGGAATGGGGGAAGATGGTGTACAGAGCTCTTCTGTGACCGGGGCA	491
OY	82	-----ValSerAs	84
Db	492	GATGTGTCTTCGTGAGACCCGACCTTTTCTTTGTCCATTGTGGCAGATGTGAGCCA	551
OY	84	PileAArgAArgPheLeuSerAlaPheHisglnProGlnValGlyLeuAlaGlnAlaAlaG	104
Db	552	CATCAGGGCGCTTCTCAGTGCATTCACAGAGCCACAGTGGGGCTCATCTACAGCGCCCA	611
OY	104	nGlnLeuLeuCyAspGlnGlnAlaProGlnAArgGlnAArgLeuAlaAspLeuLeuH	124
Db	612	GCAGCTGCTGTGTGTAGACAGAGCCCAACAGAGGACAGAGGCTGTGCTGACTCTGCA	671
OY	124	AsnValSerGlnAsnAlaAlaGlnThrArgAlaGluAspProProctrrPheglnGln	144
Db	672	CAACCTCAGCCGAACATCGCGGCCGAGACCCGGGCTGAGGACCCGCTGTGTTGAAG	731
OY	144	yleuGlnSerAArgPheGlnSerLySerGlyTyrLeuAArgTyrSerCysGlnSerArgTl	164
Db	732	CTTGAGATCCGATTTCAAGACAGCTGCTATCTAGATACAGCTGTGAGAGCGCGAT	791
OY	164	eArgSerTyrLeuAArgGluValSerSerTyrProSerThrValGlyAlaGlnAlaGlnG	184
Db	792	CCGGAGTTACCTGAGAGGAGGTAGTCTCTACCCCTCCACAGTGGGTGGAGAGCTTCAGA	851
OY	184	uGlnPheLeuAArgValLeuGlySerMetCysGlnAArgLeuAArgSerMetGlnTyrAsnG	204
	852	GGAATTCCTGGGGGTCTCTCGGCTCATGTGCCAGAGAGGCTCGGGTCCATGCATGACAATGG	911
OY	204	ySerTyrPheAspAArgGlyAlaLeuGlyGlySerAArgLeuCystrhrProGlnGlyTrpPh	224
Db	912	CAGCTACTTCGACAAAGAGCCAAAGGGGGGACCGCCCTCTCCACACCCGAAAGGCTGGTT	971
OY	224	eSerCysglnGlyProPheAspMetAspSerCysLeuSerAArgHisSerLysAsnProTy	244
Db	972	CTCTCGCCAGGGTCCCTTGACATGACAGCTGTATTAACACACTCCATCACACCCCTTA	103
OY	244	rSerAsnAArgGlnSerAArgLeuPheSerThrTrpAsnLeuAspHisIleIleGlyL	264
Db	1032	CAGTAAACGGGAGACAGAGATCCTTCAGCACCTGGAACTGTGATCACATATAGAAAA	109
OY	264	AllyAArgThrIleIleProThrLeuValGlnAlaIleLysGlnGlnAspGlyAArgGluVa	284
Db	1092	GAAACGCCACATCTCTACACTGGTGGAAACAATTAAAGGAACAAGATGGAAAGAAACT	115
OY	284	lAspTrpIuTyrrPheTyrglyLeuLeuPheThrSerGlnuAsnLeuLysLeuValHisI	304
Db	1152	GGACGTGGAGATTTTATATGGCGTCTTTTATCTCAGAGAACTTAAACTGTAGTCACAT	121

Oy		304	eValYlshStLysTstThrTHiLiShLeuSnCySaPrProSerArgIleTyrlYsPr	322
Dd		1212	tGTCtGCCATTAAGAAACACCACCACAAGCTCAACTGTGCAGCAGAGAATATTACAAACC	127
Oy		334	oGlnThrArGlLeuLysArgLySgInProValArGLySarGln	338
Dd		1272	CCAGACAAGGTTGAAGCGGAGACAGCCTGTGGGAAACGCCAG	1314
	RESULT 11			
	E36548		1038 bp	DNA linear PART 18-JUN-200
	LOCUS	Caspase-activating DNase.		
	DEFINITION	E36548		
	ACCESSION	E36548.1 GI:13022711		
	VERSION	JP 1999239494-A/1.		
	KEYWORDS	Homo sapiens.		
	SOURCE	Homo sapiens		
	ORGANISM	Homo sapiens		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	AUTHORS	Mammalia; Eutheria; Primates; Catarrhini, Homidae; Homo.		
	TITLE	1 (bases 1 to 1038)		
	JOURNAL	Shigekazu,N. and Masato,E. Caspase-activating DNase Patent: JP 1999239494-A 1 07-SEP-1999; OSAKA BIO SCI KENKYUSHO OS Homo sapiens (human) PN JP 1999239494-A/1 <u>PD 07-SEP-1999</u> PR Z5-DRC-1998 JP 1998369093		
	COMMENT			
	FEATURES			
	source	Location/Qualifiers		
		1..1038		
		/organism='Homo sapiens'		
		/db_xref='taxon:9606'		
	BASE COUNT	236 a 293 c 307 g 202 t		
	ORIGIN			
	Alignment Scores:			
	Pred. No.:	2,95e-124	Length:	1038
	Score:	1416.00	Matches:	261
	Percent Similarity:	88.17%	Conservative:	37
	Best Local Similarity:	77.22%	Mismatches:	38
	Query Match:	79.15%	Indels:	2
	DB:	6	Gaps:	1
	US-09-748-451-2 (1-338) x E36548 (1-1038)			
Oy		1	MettLeuGlnLysProLysSerValLysLeuArGalLeuArGSerProArGLysPhegly	20
Dd		10	GtGTCcCGCAACCAATAATGGCTCAAGTTGGAGGCCCTCATATAGGCCCTGCAGATTGGCG	69
Oy		21	VAlAlAgLyArGserCYsgInGLyValLeuArGLyGLyCYsLeuArGPheGLnLeuPro	40
Dd		70	GrGGGGGGCCGGAGAGCTGCCAGAGAGCTGCTGCTTAAGGGGCTGGCTGCCTTCCAAGCTCCGG	129
Oy		41	GIuArGlySerArGLeucYSLeuVrGiUaRPGLYthGCLneUthrCLnuSPrrYrPhe	60
Dd		130	ATGCCGGGTCCCggCTGCTGCTGTACGAAGATGGCACGGAGGTGACGGAGCACTGCTTC	189
Oy		61	ProSeArValProASnaAsnAlagLUleuValLleuLeuthrLeuGLyGLnAlArPrpGLngly	80
Dd		190	CCGGGccCTTCCCAACGACGAGCTAGCTCCATTATGGTCACCGCTGGCGAGACTGGCATGGC	249
Oy		81	TyrValSerAspILearGArGPheLeuSerAlaPheHisGLuProGLnValGLyleuile	100
Dd		250	TATGAGAGTGACATACACAGTTTCTCAAGTGTATTAAAGAACCAATCGCGGCGTCAATC	309
Oy		101	GIuAlAlAgInGLnLeuLeucYSapGLUGlnAlArProGLnArGLnArGLneuala	120







REFERENCE 1 (bases 1 to 155585)  
 AUTHORS Bray-Allen, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgehire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 COMMENT requests: clonerequests@sanger.ac.uk  
 On Oct 21, 2001 this sequence version replaced gi:15021313.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 9, constructed by the Sanger Centre Chromosome 9 Mapping group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr9  
 RP11-205M20 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBAC3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-205M20. It may be shorter because we sequence overlapping sections only once, except for a short overlap.  
 The true right end of clone RP11-205M20 is at 155585 in this sequence. The true left end of clone RP11-55504 is at 84490 in this sequence. The true right end of clone RP11-334P12 is at 2000 in this sequence.

FEATURES  
 source  
 1. 155585  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /clone="RP11-205M20"  
 /clone\_id="RP11-11.1"  
 BASE COUNT 48210 a 32474 c 29896 g 45005 t  
 ORIGIN

Alignment Scores:  
 No: 6.62e-106 Length: 155585  
 nt Similarity: 1251.00 Matches: 268  
 Local Similarity: 85.88% Conservative: 24  
 Mismatches: 34  
 Query Match: 78.82% Indels: 18  
 Gaps: 3

US-09-748-451-2 (1-338) x AL353671 (1-155585)  
 QY 1 MetleuGlnlyrProlyrSerValylsLeuArgAlaLeuArgSerProArglyrPhegly 20  
 DB 105080 GTGGTCCAGAGCAAGACGACGCGGCGCCCTGTCGACCCGAGAAATTTGGC 105021  
 QY 21 ValAlaGlyArgSerCysGlnGluValLeuArglyrGlyCysLeuArgPheGlnLeuPro 40  
 DB 105020 ATGGCGGGGACGAGACGCGGAGATGCTGCGCAAGGCTACCTCTGCTTACGCTCCCT 104961  
 QY 41 GluArglyrSerArgLeuGlyrGluAspGlyThrGluLeuThrGluAspTyrPhe 60  
 DB 104960 CG-----TACGAGATGGCATGAGCTGACCGAGAGC----- 104930  
 QY 61 ProSerValProAspAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGly 80

DB 104929 -----GTCCTCGACGATGCCGACCTGCTGCTCACCCTCCAGCCGCTGCAGGCC 104876  
 QY 81 TyrValSerAspIleArgArgPheLeuSerIlePheHisGlnProGlnValGlyLeuIle 100  
 DB 104875 TATGTGATGACATACGCGGTCTCTCTCAGTCCTTTTGCCACGCCGCTGCTATC 104816  
 QY 101 GlnAlaGlnGlnLeuLeuGlnCysAspGlnAla-ProGlnArgGlnArgLeuAla 120  
 DB 104815 CAGCCCGCCGCGACGCTACTGTGATGAGAGGCCGCCACAGAGGAGAGACTGTGGC 104756  
 QY 120 AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrArgAlaGluAspProPr 140  
 DB 104755 CGACCTCTGTACACGACGACCCAGACGTCGACCGCAGCCGCGCTGAGACCCGCT 104696  
 QY 140 OTrPheGlnGlyLeuGlnSerArgPheGlnSerIleGlyThrLeuArgTyrSerCys 160  
 DB 104695 GTGGTTTCAAGGCTTGGAGTCCCATTTTGAATTAAGTCTGCTATCTGAGATACACTG 104636  
 QY 160 SgluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAl 180  
 DB 104635 TGAAGCCGAGTCCGAGTACTGAGAGAGTGAAGTCTGCTCCTCATGATGGTGGTGC 104576  
 QY 180 AgluValGlnGlnGluThrLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMe 200  
 DB 104575 AGAGGCTCAGAGAGAAATTCGCGGGCTCCGCTCCATGTGCCAGAAAGCTCAGCTGT 104516  
 QY 200 TGlTyrAsnGlySerTyrPheAspArgGlyAlaGlyGlySerArgLeuGlyCysThrPr 220  
 DB 104515 GCAGTACGACGCGACGCTTAC-MACAGAGAGCCAAAGGTGGACGCTCTCTCAGACCC 104457  
 QY 220 OGluGlyTrpPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSe 240  
 DB 104456 GGAAGGCTAGTTCCTCCGAGGCTCCCTTGACATGAGCGGCTGTGTAAATATCTC 104397  
 QY 240 rIlaAspProIyrSerAsnArgGlySerArgIleLeuPheSerThrTrpAsnLeuAspH 260  
 DB 104396 CATGACCCCTACAGTAAACAGAGAGAGAGGCTCTCTTCTGACCTGGAACTGGATCA 104337  
 QY 260 sIleIleGlu-LysLysArgThrIleLeuProThrLeuValGlnAlaIleLysGlnGln 280  
 DB 104336 CATGCTGGAACAAACAAACCCGACCATTCCTACACGTGGAGAGCTTAACAGACAG 104277  
 QY 280 spGlyArgGluValAspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeu 300  
 DB 104276 ATGGAAGAGAGTGGAGCTGGGAATATTTTGTGCTCTTCTTCACTCAGAGAACTAA 104217  
 QY 300 ysluValAlaHisIleValCysHisLysLysThrThrHisLysLeuAsnCysAspProSe 320  
 DB 104216 AATTAGTGT-----GTCGCCATAGAAACACCCACCAACCAAGCTATGACCTGAACA 104161  
 QY 320 rGlyLeuTyrPProGlnThrArgLeuAsnArgLysGlnProValArgLysArgGln 338  
 DB 104160 GAATCTCAACAAACCCAGACAGAGTGAAGAGAGAGTGGCTGTGAAAGGCCGAC 104105

RESULT 16  
 LOCUS AL355811  
 DEFINITION Homo sapiens chromosome 9 clone RP11-562A11 map p13.1-21.1, \*\*\*  
 SEQUENCING IN PROGRESS \*\*\*; 26 unordered pieces.  
 ACCESSION AL355811  
 VERSION AL355811.4 GI:9863741  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_CANCELLED.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 142970)  
 AUTHORS McJannet, K.  
 TITLE Direct Submission  
 JOURNAL Submitted (09-JUL-2001) Sanger Centre, Hinxton, Cambridgehire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequests@sanger.ac.uk

## COMMENT

On Aug 21, 2000 this sequence version replaced gi:9213647.  
 ----- Genome Center  
 Center: Sanger Centre  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 ----- Project Information  
 Center project name: BA562A11  
 ----- Summary Statistics  
 Assembly program: XGAP4; version 4.5  
 Sequencing vector: plasmid; L08752; 100% of reads  
 Chemistry: Dye-terminator ET-amersham; 18% of reads Chemistry:  
 Dye-terminator Big Dye; 81% of reads  
 Consensus quality: 127694 bases at least Q40  
 Consensus quality: 133388 bases at least Q30  
 Consensus quality: 136538 bases at least Q20  
 Insert size: 140470; sum-of-contigs  
 Insert size: 155298; 6.0% error; agarose-fp  
 Quality coverage: 3.12% in Q20 bases; sum-of-contigs quality  
 coverage: 3.39% in Q20 bases; agarose-fp  
 -----  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 26 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 4033: contig of 4033 bp in length  
 4034 4133: gap of 100 bp  
 4134 8168: contig of 4035 bp in length  
 8169 8268: gap of 100 bp  
 8269 14731: contig of 6463 bp in length  
 14732 14831: gap of 100 bp  
 14832 18803: contig of 3972 bp in length  
 18804 18903: gap of 100 bp  
 18904 21145: contig of 2242 bp in length  
 21146 21245: gap of 100 bp  
 21246 26247: contig of 5002 bp in length  
 26248 26347: gap of 100 bp  
 26348 30120: contig of 3773 bp in length  
 30121 30220: gap of 100 bp  
 30221 32361: contig of 2241 bp in length  
 32362 32561: gap of 100 bp  
 32562 33911: contig of 7050 bp in length  
 33912 39711: gap of 100 bp  
 39712 48337: contig of 8628 bp in length  
 48338 48437: gap of 100 bp  
 48438 50511: contig of 2074 bp in length  
 50512 50611: gap of 100 bp  
 50612 52986: contig of 2375 bp in length  
 52987 53086: gap of 100 bp  
 53087 59466: contig of 6380 bp in length  
 59467 59566: gap of 100 bp  
 59567 64909: contig of 5343 bp in length  
 64910 65009: gap of 100 bp  
 65010 69805: contig of 4796 bp in length  
 69806 69905: gap of 100 bp  
 69906 75879: contig of 5974 bp in length  
 75880 75979: gap of 100 bp  
 75980 81017: contig of 5038 bp in length  
 81018 81117: gap of 100 bp  
 81118 95015: contig of 13988 bp in length  
 95016 95115: gap of 100 bp  
 95116 99680: contig of 4565 bp in length  
 99681 99780: gap of 100 bp  
 99781 104130: contig of 4350 bp in length  
 104131 104230: gap of 100 bp  
 104231 110144: contig of 5914 bp in length  
 110145 110244: gap of 100 bp  
 110245 115096: contig of 4852 bp in length  
 115097 115196: gap of 100 bp

FEATURES  
source

\* 115197 118257: contig of 3061 bp in length  
 \* 118258 118357: gap of 100 bp  
 \* 118358 126637: contig of 8280 bp in length  
 \* 126638 126737: gap of 100 bp  
 \* 126738 132827: contig of 6090 bp in length  
 \* 132828 132927: gap of 100 bp  
 \* 132928 142970: contig of 10043 bp in length.  
 Location/Qualifiers  
 1. 142970  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="9"  
 /map="p13.1-21.1"  
 /clone="RP11-562A11"  
 /clone\_1hd="RCR11.2"  
 1. 4033  
 /note="assembly-fragment:01252  
 clone\_end:SP6  
 vector\_side:left"  
 4134. 8168  
 /note="assembly-fragment:00004  
 fragment\_chain:1"  
 8269. 14731  
 /note="assembly-fragment:01032  
 fragment\_chain:1"  
 14832. 18803  
 /note="assembly-fragment:00497  
 fragment\_chain:2"  
 18904. 21145  
 /note="assembly-fragment:00407  
 fragment\_chain:2"  
 21246. 26247  
 /note="assembly-fragment:01246  
 fragment\_chain:3"  
 26348. 30120  
 /note="assembly-fragment:00027  
 fragment\_chain:3"  
 30221. 32361  
 /note="assembly-fragment:00019"  
 32362. 32561  
 /note="assembly-fragment:00023"  
 32562. 33911  
 /note="assembly-fragment:00055"  
 33912. 39711  
 /note="assembly-fragment:00090"  
 39712. 48337  
 /note="assembly-fragment:00183"  
 48338. 48437  
 /note="assembly-fragment:00202"  
 48438. 50511  
 /note="assembly-fragment:00332"  
 50512. 50611  
 /note="assembly-fragment:00342"  
 50612. 52986  
 /note="assembly-fragment:00567"  
 52987. 53086  
 /note="assembly-fragment:00580"  
 53087. 59466  
 /note="assembly-fragment:00586"  
 59467. 59566  
 /note="assembly-fragment:00596"  
 59567. 64909  
 /note="assembly-fragment:00605"  
 64910. 65009  
 /note="assembly-fragment:00615"  
 65010. 69805  
 /note="assembly-fragment:00625"  
 69806. 69905  
 /note="assembly-fragment:00635"  
 69906. 75879  
 /note="assembly-fragment:00645"  
 75880. 75979  
 /note="assembly-fragment:00655"  
 75980. 81017  
 /note="assembly-fragment:00665"  
 81018. 81117  
 /note="assembly-fragment:00675"  
 81118. 95015  
 /note="assembly-fragment:00685"  
 95016. 95115  
 /note="assembly-fragment:00695"  
 95116. 99680  
 /note="assembly-fragment:00705"  
 99681. 99780  
 /note="assembly-fragment:00715"  
 99781. 104130  
 /note="assembly-fragment:00725"  
 104131. 104230  
 /note="assembly-fragment:00735"  
 104231. 110144  
 /note="assembly-fragment:00745"  
 110145. 110244  
 /note="assembly-fragment:00755"  
 110245. 115096  
 /note="assembly-fragment:00765"  
 115097. 115196  
 /note="assembly-fragment:00775"  
 115197. 118257  
 /note="assembly-fragment:00785"  
 118258. 126637  
 /note="assembly-fragment:00795"  
 126638. 126737  
 /note="assembly-fragment:00805"  
 126738. 132827  
 /note="assembly-fragment:00815"  
 132828. 132927  
 /note="assembly-fragment:00825"  
 132928. 142970  
 /note="assembly-fragment:00835"  
 142971. 142970  
 /note="assembly-fragment:00845"  
 142971. 142970  
 /note="assembly-fragment:00855"  
 142971. 142970  
 /note="assembly-fragment:00865"  
 142971. 142970  
 /note="assembly-fragment:00875"  
 142971. 142970  
 /note="assembly-fragment:00885"  
 142971. 142970  
 /note="assembly-fragment:00895"  
 142971. 142970  
 /note="assembly-fragment:00905"  
 142971. 142970  
 /note="assembly-fragment:00915"  
 142971. 142970  
 /note="assembly-fragment:00925"  
 142971. 142970  
 /note="assembly-fragment:00935"  
 142971. 142970  
 /note="assembly-fragment:00945"  
 142971. 142970  
 /note="assembly-fragment:00955"  
 142971. 142970  
 /note="assembly-fragment:00965"  
 142971. 142970  
 /note="assembly-fragment:00975"  
 142971. 142970  
 /note="assembly-fragment:00985"  
 142971. 142970  
 /note="assembly-fragment:00995"  
 142971. 142970  
 /note="assembly-fragment:01005"  
 142971. 142970  
 /note="assembly-fragment:01015"  
 142971. 142970  
 /note="assembly-fragment:01025"  
 142971. 142970  
 /note="assembly-fragment:01035"  
 142971. 142970  
 /note="assembly-fragment:01045"  
 142971. 142970  
 /note="assembly-fragment:01055"  
 142971. 142970  
 /note="assembly-fragment:01065"  
 142971. 142970  
 /note="assembly-fragment:01075"  
 142971. 142970  
 /note="assembly-fragment:01085"  
 142971. 142970  
 /note="assembly-fragment:01095"  
 142971. 142970  
 /note="assembly-fragment:01105"  
 142971. 142970  
 /note="assembly-fragment:01115"  
 142971. 142970  
 /note="assembly-fragment:01125"  
 142971. 142970  
 /note="assembly-fragment:01135"  
 142971. 142970  
 /note="assembly-fragment:01145"  
 142971. 142970  
 /note="assembly-fragment:01155"  
 142971. 142970  
 /note="assembly-fragment:01165"  
 142971. 142970  
 /note="assembly-fragment:01175"  
 142971. 142970  
 /note="assembly-fragment:01185"  
 142971. 142970  
 /note="assembly-fragment:01195"  
 142971. 142970  
 /note="assembly-fragment:01205"  
 142971. 142970  
 /note="assembly-fragment:01215"  
 142971. 142970  
 /note="assembly-fragment:01225"  
 142971. 142970  
 /note="assembly-fragment:01235"  
 142971. 142970  
 /note="assembly-fragment:01245"  
 142971. 142970  
 /note="assembly-fragment:01255"  
 142971. 142970  
 /note="assembly-fragment:01265"  
 142971. 142970  
 /note="assembly-fragment:01275"  
 142971. 142970  
 /note="assembly-fragment:01285"  
 142971. 142970  
 /note="assembly-fragment:01295"  
 142971. 142970  
 /note="assembly-fragment:01305"  
 142971. 142970  
 /note="assembly-fragment:01315"  
 142971. 142970  
 /note="assembly-fragment:01325"  
 142971. 142970  
 /note="assembly-fragment:01335"  
 142971. 142970  
 /note="assembly-fragment:01345"  
 142971. 142970  
 /note="assembly-fragment:01355"  
 142971. 142970  
 /note="assembly-fragment:01365"  
 142971. 142970  
 /note="assembly-fragment:01375"  
 142971. 142970  
 /note="assembly-fragment:01385"  
 142971. 142970  
 /note="assembly-fragment:01395"  
 142971. 142970  
 /note="assembly-fragment:01405"  
 142971. 142970  
 /note="assembly-fragment:01415"  
 142971. 142970  
 /note="assembly-fragment:01425"  
 142971. 142970  
 /note="assembly-fragment:01435"  
 142971. 142970  
 /note="assembly-fragment:01445"  
 142971. 142970  
 /note="assembly-fragment:01455"  
 142971. 142970  
 /note="assembly-fragment:01465"  
 142971. 142970  
 /note="assembly-fragment:01475"  
 142971. 142970  
 /note="assembly-fragment:01485"  
 142971. 142970  
 /note="assembly-fragment:01495"  
 142971. 142970  
 /note="assembly-fragment:01505"  
 142971. 142970  
 /note="assembly-fragment:01515"  
 142971. 142970  
 /note="assembly-fragment:01525"  
 142971. 142970  
 /note="assembly-fragment:01535"  
 142971. 142970  
 /note="assembly-fragment:01545"  
 142971. 142970  
 /note="assembly-fragment:01555"  
 142971. 142970  
 /note="assembly-fragment:01565"  
 142971. 142970  
 /note="assembly-fragment:01575"  
 142971. 142970  
 /note="assembly-fragment:01585"  
 142971. 142970  
 /note="assembly-fragment:01595"  
 142971. 142970  
 /note="assembly-fragment:01605"  
 142971. 142970  
 /note="assembly-fragment:01615"  
 142971. 142970  
 /note="assembly-fragment:01625"  
 142971. 142970  
 /note="assembly-fragment:01635"  
 142971. 142970  
 /note="assembly-fragment:01645"  
 142971. 142970  
 /note="assembly-fragment:01655"  
 142971. 142970  
 /note="assembly-fragment:01665"  
 142971. 142970  
 /note="assembly-fragment:01675"  
 142971. 142970  
 /note="assembly-fragment:01685"  
 142971. 142970  
 /note="assembly-fragment:01695"  
 142971. 142970  
 /note="assembly-fragment:01705"  
 142971. 142970  
 /note="assembly-fragment:01715"  
 142971. 142970  
 /note="assembly-fragment:01725"  
 142971. 142970  
 /note="assembly-fragment:01735"  
 142971. 142970  
 /note="assembly-fragment:01745"  
 142971. 142970  
 /note="assembly-fragment:01755"  
 142971. 142970  
 /note="assembly-fragment:01765"  
 142971. 142970  
 /note="assembly-fragment:01775"  
 142971. 142970  
 /note="assembly-fragment:01785"  
 142971. 142970  
 /note="assembly-fragment:01795"  
 142971. 142970  
 /note="assembly-fragment:01805"  
 142971. 142970  
 /note="assembly-fragment:01815"  
 142971. 142970  
 /note="assembly-fragment:01825"  
 142971. 142970  
 /note="assembly-fragment:01835"  
 142971. 142970  
 /note="assembly-fragment:01845"  
 142971. 142970  
 /note="assembly-fragment:01855"  
 142971. 142970  
 /note="assembly-fragment:01865"  
 142971. 142970  
 /note="assembly-fragment:01875"  
 142971. 142970  
 /note="assembly-fragment:01885"  
 142971. 142970  
 /note="assembly-fragment:01895"  
 142971. 142970  
 /note="assembly-fragment:01905"  
 142971. 142970  
 /note="assembly-fragment:01915"  
 142971. 142970  
 /note="assembly-fragment:01925"  
 142971. 142970  
 /note="assembly-fragment:01935"  
 142971. 142970  
 /note="assembly-fragment:01945"  
 142971. 142970  
 /note="assembly-fragment:01955"  
 142971. 142970  
 /note="assembly-fragment:01965"  
 142971. 142970  
 /note="assembly-fragment:01975"  
 142971. 142970  
 /note="assembly-fragment:01985"  
 142971. 142970  
 /note="assembly-fragment:01995"  
 142971. 142970  
 /note="assembly-fragment:02005"  
 142971. 142970  
 /note="assembly-fragment:02015"  
 142971. 142970  
 /note="assembly-fragment:02025"  
 142971. 142970  
 /note="assembly-fragment:02035"  
 142971. 142970  
 /note="assembly-fragment:02045"  
 142971. 142970  
 /note="assembly-fragment:02055"  
 142971. 142970  
 /note="assembly-fragment:02065"  
 142971. 142970  
 /note="assembly-fragment:02075"  
 142971. 142970  
 /note="assembly-fragment:02085"  
 142971. 142970  
 /note="assembly-fragment:02095"  
 142971. 142970  
 /note="assembly-fragment:02105"  
 142971. 142970  
 /note="assembly-fragment:02115"  
 142971. 142970  
 /note="assembly-fragment:02125"  
 142971. 142970  
 /note="assembly-fragment:02135"  
 142971. 142970  
 /note="assembly-fragment:02145"  
 142971. 142970  
 /note="assembly-fragment:02155"  
 142971. 142970  
 /note="assembly-fragment:02165"  
 142971. 142970  
 /note="assembly-fragment:02175"  
 142971. 142970  
 /note="assembly-fragment:02185"  
 142971. 142970  
 /note="assembly-fragment:02195"  
 142971. 142970  
 /note="assembly-fragment:02205"  
 142971. 142970  
 /note="assembly-fragment:02215"  
 142971. 142970  
 /note="assembly-fragment:02225"  
 142971. 142970  
 /note="assembly-fragment:02235"  
 142971. 142970  
 /note="assembly-fragment:02245"  
 142971. 142970  
 /note="assembly-fragment:02255"  
 142971. 142970  
 /note="assembly-fragment:02265"  
 142971. 142970  
 /note="assembly-fragment:02275"  
 142971. 142970  
 /note="assembly-fragment:02285"  
 142971. 142970  
 /note="assembly-fragment:02295"  
 142971. 142970  
 /note="assembly-fragment:02305"  
 142971. 142970  
 /note="assembly-fragment:02315"  
 142971. 142970  
 /note="assembly-fragment:02325"  
 142971. 142970  
 /note="assembly-fragment:02335"  
 142971. 142970  
 /note="assembly-fragment:02345"  
 142971. 142970  
 /note="assembly-fragment:02355"  
 142971. 142970  
 /note="assembly-fragment:02365"  
 142971. 142970  
 /note="assembly-fragment:02375"  
 142971. 142970  
 /note="assembly-fragment:02385"  
 142971. 142970  
 /note="assembly-fragment:02395"  
 142971. 142970  
 /note="assembly-fragment:02405"  
 142971. 142970  
 /note="assembly-fragment:02415"  
 142971. 142970  
 /note="assembly-fragment:02425"  
 142971. 142970  
 /note="assembly-fragment:02435"  
 142971. 142970  
 /note="assembly-fragment:02445"  
 142971. 142970  
 /note="assembly-fragment:02455"  
 142971. 142970  
 /note="assembly-fragment:02465"  
 142971. 142970  
 /note="assembly-fragment:02475"  
 142971. 142970  
 /note="assembly-fragment:02485"  
 142971. 142970  
 /note="assembly-fragment:02495"  
 142971. 142970  
 /note="assembly-fragment:02505"  
 142971. 142970  
 /note="assembly-fragment:02515"  
 142971. 142970  
 /note="assembly-fragment:02525"  
 142971. 142970  
 /note="assembly-fragment:02535"  
 142971. 142970  
 /note="assembly-fragment:02545"  
 142971. 142970  
 /note="assembly-fragment:02555"  
 142971. 142970  
 /note="assembly-fragment:02565"  
 142971. 142970  
 /note="assembly-fragment:02575"  
 142971. 142970  
 /note="assembly-fragment:02585"  
 142971. 142970  
 /note="assembly-fragment:02595"  
 142971. 142970  
 /note="assembly-fragment:02605"  
 142971. 142970  
 /note="assembly-fragment:02615"  
 142971. 142970  
 /note="assembly-fragment:02625"  
 142971. 142970  
 /note="assembly-fragment:02635"  
 142971. 142970  
 /note="assembly-fragment:02645"  
 142971. 142970  
 /note="assembly-fragment:02655"  
 142971. 142970  
 /note="assembly-fragment:02665"  
 142971. 142970  
 /note="assembly-fragment:02675"  
 142971. 142970  
 /note="assembly-fragment:02685"  
 142971. 142970  
 /note="assembly-fragment:02695"  
 142971. 142970  
 /note="assembly-fragment:02705"  
 142971. 142970  
 /note="assembly-fragment:02715"  
 142971. 142970  
 /note="assembly-fragment:02725"  
 142971. 142970  
 /note="assembly-fragment:02735"  
 142971. 142970  
 /note="assembly-fragment:02745"  
 142971. 142970  
 /note="assembly-fragment:02755"  
 142971. 142970  
 /note="assembly-fragment:02765"  
 142971. 142970  
 /note="assembly-fragment:02775"  
 142971. 142970  
 /note="assembly-fragment:02785"  
 142971. 142970  
 /note="assembly-fragment:02795"  
 142971. 142970  
 /note="assembly-fragment:02805"  
 142971. 142970  
 /note="assembly-fragment:02815"  
 142971. 142970  
 /note="assembly-fragment:02825"  
 142971. 142970  
 /note="assembly-fragment:02835"  
 142971. 142970  
 /note="assembly-fragment:02845"  
 142971. 142970  
 /note="assembly-fragment:02855"  
 142971. 142970  
 /note="assembly-fragment:02865"  
 142971. 142970  
 /note="assembly-fragment:02875"  
 142971. 142970  
 /note="assembly-fragment:02885"  
 142971. 142970  
 /note="assembly-fragment:02895"  
 142971. 142970  
 /note="assembly-fragment:02905"  
 142971. 142970  
 /note="assembly-fragment:02915"  
 142971. 142970  
 /note="assembly-fragment:02925"  
 142971. 142970  
 /note="assembly-fragment:02935"  
 142971. 142970  
 /note="assembly-fragment:02945"  
 142971. 142970  
 /note="assembly-fragment:02955"  
 142971. 142970  
 /note="assembly-fragment:02965"  
 142971. 142970  
 /note="assembly-fragment:02975"  
 142971. 142970  
 /note="assembly-fragment:02985"  
 142971. 142970  
 /note="assembly-fragment:02995"  
 142971. 142970  
 /note="assembly-fragment:03005"  
 142971. 142970  
 /note="assembly-fragment:03015"  
 142971. 142970  
 /note="assembly-fragment:03025"  
 142971. 142970  
 /note="assembly-fragment:03035"  
 142971. 142970  
 /note="assembly-fragment:03045"  
 142971. 142970  
 /note="assembly-fragment:03055"  
 142971. 142970  
 /note="assembly-fragment:03065"  
 142971. 142970  
 /note="assembly-fragment:03075"  
 142971. 142970  
 /note="assembly-fragment:03085"  
 142971. 142970  
 /note="assembly-fragment:03095"  
 142971. 142970  
 /note="assembly-fragment:03105"  
 142971. 142970  
 /note="assembly-fragment:03115"  
 142971. 142970  
 /note="assembly-fragment:03125"  
 142971. 142970  
 /note="assembly-fragment:03135"  
 142971. 142970  
 /note="assembly-fragment:03145"  
 142971. 142970  
 /note="assembly-fragment:03155"  
 142971. 142970  
 /note="assembly-fragment:03165"  
 142971. 142970  
 /note="assembly-fragment:03175"  
 142971. 142970  
 /note="assembly-fragment:03185"  
 142971. 142970  
 /note="assembly-fragment:03195"  
 142971. 142970  
 /note="assembly-fragment:03205"  
 142971. 142970  
 /note="assembly-fragment:03215"  
 142971. 142970  
 /note="assembly-fragment:03225"

misc_feature	132928..142970	/note="assembly_fragment:01304
BASE COUNT	40736 a 27602 c 28913 g 43167 t	2552 others
ORIGIN	vector_side:right"	
Alignment Scores:	clone_end:T7	
Pred. No.:	2.07e-88	Length: 142970
Score:	1065.00	Matches: 227
Percent Similarity:	87.50%	Conservative: 18
Best Local Similarity:	81.07%	Mismatches: 24
Query Match:	59.53%	Indels: 14
DB:	2	Gaps: 2
US-09-748-451-2 (1-338) x AL355811 (1-142970)		
QY	62	SeValBrpAaSpAaNaLaGlueValLeuLeuThrlLeuGlYglnAlaTrpGlnGlyTr 81
	65024	AGCCTCTCGACGATGCCAGATCGATGCTGCTCCTCAGCCACAGCCAGGCTGGAGGCTAT 65083
	82	ValSerAaPrlLeaTgArpHeLeuSerAlaPheHisGluPrGlnValGlyLeuGln 101
	65084	GTCGATACATCGGGCTCTCCACAGCTTTTCG-----CCG 65122
QY	102	AlaAlaGlnGlnLeuLeuCysAaPglGlnAla-ProGlnAaGlnAaLeu-LeuAla 121
	65123	CC-GCCGGGACGCTACCTGTGTATGACACAGCCCCACAGGACAGACTGCTGGCCG 65181
QY	121	SpLeuLeuHHisAaPnaValSerGlnAaPnLeaAlaGlnThrArGaLaGlnAaPProT 141
	65182	ACCTCCGTGACACAGTGCACAAACAGCTCCACCCACACCCGGCTGAGGACCCGCTGT 65241
QY	141	rPheGlnGlnGlnGlnSerAaPheGlnSerGlyTrpLeuAaGlyTrpSerCysG 161
	65242	GATTGAAGGCTTGGAATCCCGATTTGGAAATGCTGGCTATGAGATACAGCTGTG 65301
QY	161	LeuSerAaGlnLeaGSerTrpLeuAaGlnValSerSerTrpProSerThrValGlyAlaG 181
	65302	AAAGCCGAGATCCGAGTATACCTGAGAGAGGTGACGTCCTGCTCCATGCTGGGTGCGAG 65361
QY	181	LeuAlaGlnGlnGlnPheLeuAaGValLeuGlySerMetCysGlnAaGlnAaGSerMetG 201
	65362	AGGCTCGAGGAGATTCCTCGGGCTCTCCGCTCCATGTCGCCGAAGCTCCAGCTGTGTC 65421
QY	201	LeuTrpAaGlnSerTrpPheAaPArGlnValAlaGlyGlnSerAaGlnCysTrpProG 221
	65422	AGTACGACGCGACGATATAC-AACAGAGAGACCAGAGGTGGACGCCCTCTGTGACACCGG 65480
QY	221	LeuGlyTrpPheSerCysGlnGlyProPheAaPmetAaSerCysLeuSerAaGlnHisSerI 241
	65481	AAGGCTAGTTCCTCGTCCGACGAGGCCCTTCGACATGAGAGCGGTGCTTGTCAAGATCTCCA 65540
	241	LeuAaPProTTrpSerAaPnaGlnSerAaGlnLeuPheSerThrTrpAaPnaPnaHisI 261
	65541	TCAACCCCTACAGTAAAGGGAAGAGAGAGGCTCTCTTCAGAGACCTGGAACCTGGATCACA 65600
QY	261	LeuLeuGln-LysLysAaTrpThrlLeuProThrlLeuValGlnAlaLeuGlnGlnAaP 280
	65601	TAGTGGAAACAAAAAGCCACCATCATCTTCACACTGTGGAGACGACTTAAAGAACAGAT 65660
QY	281	GlyAaGlnValAaPTrpGlnTrpPheTrpGlyLeuLeuPheThrsSerGlnAaPnaLeuLys 300
	65661	GGAAGAGAAAGTGCACGTGGAAATTTTATATGGCTGCTTTTCTACCTCAAGAACCTGAATAA 65720
QY	301	LeuValHisHisLeuValCysHisLysLysTrpThrlHisLysLeuAaPnaCysAaPProSerAaG 320
	65721	TTAGTGT---GTCTGCCATTAAGAAACACCCACAAACCAAGCATATGACTGAACAGA 65776
QY	321	LeuTrpLysProGlnThrArGaLeuLysAaGlyGlnProValArgLysAaGln 338
	65777	ATCTACAAACCCACAGACAGGTTGAAGAGAAAGTGGCTGTGTGTGAAGAGGCGAG 65830

LOCUS	AF406761	1499 bp	mRNA	linear	VRT 02-DEC-2001
DEFINITION	Gallus gallus caspase-activated DNase (CAD) mRNA, complete cds.				
ACCESSION	AF406761				
VERSION	AF406761.1	GI:17226273			
KEYWORDS					
SOURCE	Gallus gallus.				
ORGANISM	Gallus gallus				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.				
AUTHORS	1 (bases 1 to 1499)				
TITLE	Samejima, K., Tone, S. and Earnshaw, W.C.				
JOURNAL	CAD/DEP40 nuclease is dispensable for high molecular weight DNA				
MEDLINE	cleavage and stage I chromatin condensation in apoptosis				
REFERENCE	J. Biol. Chem. 276 (48), 45427-45432 (2001)				
AUTHORS	2 (bases 1 to 1499)				
TITLE	Samejima, K.				
JOURNAL	Direct Submission				
MEDLINE	Submitted (06-AUG-2001) ICMR, University of Edinburgh, Mayfield				
REFERENCE	Road, Edinburgh EH9 3JR, UK				
FEATURES	Location/Qualifiers				
source	1..1499				
gene	/organism="Gallus gallus"				
CDS	/db_xref="taxon:9031"				
	1..1499				
	/gene="CAD"				
	1..1002				
	/gene="CAD"				
	/note="nuclease"				
	/codon_start=1				
	/product="caspase-activated DNase"				
	/protein_id="AAL37716.1"				
	/db_xref="GI:17226274"				
	/translation="MAAPLRFRRLRPPSAQKFGAAGSLNGLRKRCRLQLPLAGS				
	RLCIYEDGETELSEAFRTLPOTLETIVLLRPESWPGCCGDIYERFLAALSRDVAVEE				
	ARRLIEDERARRQRLADLIHLNLSGSAARBDGMEFGLERFNKSSMYVSC				
	ESRIISYKWEKRSNFIISNHPAPADAKRIILDMSPDKLSVYNGSEFPTREDEAMR				
	CTKGWFCGCGPDRDDCPCKSGINPTNRRSRLIFSWNLDHITTEKKRSVPLEAE				
	VKTDEGFRVMEYFVOLLFTMMNKLIVHIACHKKTNNHLSCDKTKIYKRRQTHIIS				
BASE COUNT	402 a	345 c	403 g	349 t	
ORIGIN					
Alignment Scores:					
Pred. No.:	7,15e-89	Length:	1499		
Score:	1043.00	Matches:	197		
Percent Similarity:	75.61%	Conservative:	51		
Best Local Similarity:	60.06%	Mismatches:	74		
Query Match:	58.30%	Indels:	6		
DB:	5	Gaps:	2		
IDS-09-748-451-2 (1-338) x AF406761 (1-1499)					
QY	6	lysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGlyArgSer	25		
Db	16	AGGGGGTCCCGCTCCCGCCGCCGCCGCGAGAGTTCGGGGCGCGCGAGCC	75		
QY	26	CysGlnGluValAlaLeuArgLysGlyCysLeuArgPheGlnLeuProGluArgLysArg	45		
Db	76	CTGGCTGGGGCTGCTGGGAGAGGGCTGCTGAGGCTCCGTTGGCAGGACCGG	135		
QY	46	LeuGlyLeuArgLysPheGlyValAlaGlyArgSerProArgLysPheGlyValAlaGlyArgSer	65		
Db	136	CTGGCTGGGGCTGCTGGGAGAGGGCTGCTGAGGCTCCGTTGGCAGGACCGG	195		
QY	66	AsnAlaGluLeuValAlaLeuArgLysGlyCysLeuArgPheGlnLeuProGluArgLysArg	85		
Db	196	CAGACGGAGCTGGGCTGCTGCGCGCGCGGAGAGCTGCGCGGCTGCGGGGACCTG	255		
QY	86	ArgArgPheLeuSerAlaPheHisGluProGluValGlyLeuIleGlnAlaGlnGln	105		

```

Db      256 GAGCGTCTCTGCGCCGCTGTGACGCGGACGCGCGCGTGTGTGAGCGCCGCGCGCGG 315
      |||||||.....|
Qy      106 LeuLeuCysAspGluGlnAlaProGlnArgGlnLeuLeuAlaAspLeuLeuHisAsn 125
      |||||||.....|
Db      316 CTCTGTGAGAGAGAGCGCCGCGGACGAGAGGCTGTGTGGGATCTCTCCACAAAC 375
      |||||||.....|
Qy      126 ValSerGlnAsnIleAlaIleGluThrArgAlaGluAspProTrophegIuLeu 145
      |||||||.....|
Db      376 CTGAGCGGGGACGCGCGCGGACGCGGACGAGAGCGGAGAGGTTCTCAGGGCTTC 435
      |||||||.....|
Qy      146 GluSerArgPheGlnSerLysSerGlyThrLeuArgTyrSerCysGluSerArgIleArg 165
      |||||||.....|
Db      436 GAATCTCGTTTAAACAATAATCAGCTATATCGCTACAGCTGTAAGCGAATACGAA 495
      |||||||.....|
Qy      166 SerTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGluAlaGlnGlu 185
      |||||||.....|
Db      496 AGCTACATGAAGAGCTTAAATTTATTTCAAATGTCATCAGCTACAGCGAGATGCA 555
      |||||||.....|
Qy      186 PheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrAsnGlySer 205
      |||||||.....|
Db      556 TATAAAGGATACCTTGACCTGATGTGCGATTAAGTGAATCTGTGAATATACGCGCAGC 615
      |||||||.....|
Qy      206 TyrPheAspArg---GlyAlaIleGlySerArgLeuCysThrProGlnGlyTyrPhe 224
      |||||||.....|
Db      616 TACTTTGACAGACAGACAGAGAGCGCGGATGCGCTGTGACAGAGAGGATGGTTC 675
      |||||||.....|
Qy      225 SerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsnProTyr 244
      |||||||.....|
Db      676 TCTGTGACGCGCGCTTTTGACAGAGATGACGCGCATGTAAAGATCTATACACCCCTAC 735
      |||||||.....|
Qy      245 SerAsnArgGluSerArgIleLeuPheSerThrTyrPheAsnAspHisIleIleGluLys 264
      |||||||.....|
Db      736 ACTAACAGGAGAAAGAGAGATCTCTTACAGCATGTGAGATCTCATCTAATTAAGAAAG 795
      |||||||.....|
Qy      265 LysArgThrIleIleProThrLeuValGluAlaIleLysGlnGlnAspIleArgIleVal 284
      |||||||.....|
Db      796 AAACGACGAGTGTGTCCAGAGACTGCGAGAGCTGTCAAAACAGAGATGAGAGAAAGTGC 855
      |||||||.....|
Qy      285 AspTyrGluTyrPheTyrGlyLeuLeuPheThrSerGlnAsnLeuLysLeuValHisIle 304
      |||||||.....|
Db      856 AACGCGAGTACTTTTAAACCTATCTTGTACAAATGAATATCTAAACTTGTACATAT 915
      |||||||.....|
Qy      305 ValCysHisLysLysThrThrHisLysLeuAsnCysAspProSerArgIleTyrLysPro 324
      |||||||.....|
Db      916 GCTGCCACAAAGAAACCAATCATATCTCAGCTGTGAGAAACCAATTAATTCAGCG--- 972
      |||||||.....|
Qy      325 GlnThrArgLeuLysArgLysGln 332
      |||||||.....|
Db      973 -----AAAAGGAGCA 984
      |||||||.....|

RESULT 18
AF286179 1159 bp mRNA linear VRT 15-AUG-2000
DEFINITION Dantio relio caspase-activated Dnase mRNA, complete cds.
ACCESSION AF286179
VERSION AF286179.1 GI:9802391
KEYWORDS
SOURCE
ORGANISM
Dantio relio.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Dantio.
REFERENCE
1 (bases 1 to 1159)
AUTHORS Liu,C.C. and Huang,C.J.
TITLE Direct Submission
JOURNAL Submitted (10-JUL-2000) Institute of Biological Chemistry, Academia
Sinica, 128, Yen-Chu-Yuan Road, Section 2, Taipei 115, Taiwan
FEATURES
source
1..1159
/organism="Dantio relio"
/db_xref="taxon:7955"
25..1026
CDS

```

```

/codon_start=1
/product="caspase-activated Dnase"
/protein_id="AA099706.1"
/db_xref="gi:9802392"
/translation="MKTTRKLVKIRISANQARKVITAANVKEIKRKCOLLRVSS
GVLVCLYEDGVTVEEYFQNLIPDWTDLVLPQDSNNGPADININVLIDNTELLY
SAAQGLSDERSPKRRRLIDGLSNLNRSEKRYKVALEADHDFEGIDYFPRKSAVMY
NCSRIKGLYKLVDDGYTQITINAKTSEYKVALEALKAARIDYFPRSEKDN
RLCTEGRWFSKCGADENCFNLHINPYGNRESRIIFSTNIDHLIERKRTVIPALA
KALEANKSNDIVDYFYRLITRENKLVITVCHRKGAHELSDSRKIKYKRVKTV"
BASE COUNT 370 a 240 c 275 g 274 t
ORIGIN

```

```

Alignment Scores:
Pred. No.: 2,14e-79 Length: 1159
Score: 941.00 Matches: 169
Percent Similarity: 71.95% Conservative: 67
Best Local Similarity: 51.52% Mismatches: 88
Query Match: 52.60% Indels: 4
DB: Gaps: 2
US-09-748-451-2 (1-338) x AF286179 (1-1159)

```

```

Qy      4 LysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGly 23
      |||||||.....|
Db      40 AAACCCAACTGGTTAGATTAAGAAAGCCCAACAGCGCTGAATAATACGAATTCGCGCT 99
      |||||||.....|
Qy      24 ArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuProGlnArgLys 43
      |||||||.....|
Db      100 GTTAACTTAAGAGAGGTGATTAATAAAGAGATGTCAATTAAGAGATTCACAGCTCCGT 159
      |||||||.....|
Qy      44 SerArgLysLeuTyrGlnLysPheGlnLeuArgPheGlnLeuArgPheProSerVal 63
      |||||||.....|
Db      160 GTACGCTGCTGCTTAATACAGAGATGAGAGTACTGTACCAAGATTAATTCACAAATCTT 219
      |||||||.....|
Qy      64 ProAspAsnIleGlnLeuValLeuLeuThrLeuGlnAlaIleArgPheGlnIleTyrValSer 83
      |||||||.....|
Db      220 CCGTATTAACACAGATTAATTAATTCCTGCGCCATGAGACAGAGCTGAATGGTTTCAGAT 279
      |||||||.....|
Qy      84 AspIleArgArgPheLeuSerAlaPheHisGlnProGlnValGlyLeuIleGlnIleAla 103
      |||||||.....|
Db      280 GAATTAATGAGAGTGGTGGCTGACAGAAACACAGAGGAGCGTGGTTCAGTCCAGCA 339
      |||||||.....|
Qy      104 GlnGlnLeuLeuGlyAspGlnGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeu 123
      |||||||.....|
Db      340 CAAGCGCTTGTGCATATGATACGTTCCGCCCAAGCGCGAGATCTGGAGANCTGCTG 399
      |||||||.....|
Qy      124 HisAsnValSerGlnAsnIleAlaIleGluThrArgAlaGluAspProTrophegIu 143
      |||||||.....|
Db      400 AGTAACCTGAGGAGACCGTTCAGAGCTGAGAAACAGACTACAGATCATGACTGTGAA 459
      |||||||.....|
Qy      144 GlyLeuGlnSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerArg 163
      |||||||.....|
Db      460 GGTATTTGATACCGCGCTTTAAACCAAGTCTGCTTACATGAAGTACAAATTCGAGAGAGA 519
      |||||||.....|
Qy      164 IleArgSerTyrLeuArgGluValSerSerTyrProSerThrVal---GlyAlaGluAla 182
      |||||||.....|
Db      520 ATCCGCGGATATTTAAGAGAGGTGATGATGATCTCAAACTATTAATAATCCAAACCA 579
      |||||||.....|
Qy      183 GlnGlnGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyr 202
      |||||||.....|
Db      580 AAAAGTAAATATTAAGAGAGCGCGAGCGCTGCTGCTGAGAACTGAAGCGTGGAGTAC 639
      |||||||.....|
Qy      203 AsnGlySerTyrPheAspArgGlyAlaLysGlyLysSerArgLysArgTyrProGlnGly 222
      |||||||.....|
Db      640 AACGCGACCTATTTTACAGATCAAGAGAGAGCGCAACAGACTGCTGCACTGAAGAGGCG 699
      |||||||.....|
Qy      223 TrpPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsn 242
      |||||||.....|
Db      700 TGGTTCCTGCGCAGGAGCTCTTTGATAGAAACGCTGCANTTTCTGATTCATTAAC 759
      |||||||.....|
Qy      243 ProTyrSerAsnArgLysSerArgIleLeuPheSerThrTyrPheAsnAspHisIleIle 262
      |||||||.....|

```

RESULT	19	AL691523	160716 bp	DNA, linear	HTG_12-AUG-2002
SEQUENCE					
CLONING					
DESCRIPTION					
ACCESSION		AL691523			
VERSION		AL691523.5	GI:21955581		
KEYWORDS		HTG; HTGS_PHAZE2; HTGS_ACTIVETERIN; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE		human.			
ORGANISM		Homo sapiens			

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
Direct Submission  
White, S.  
1 (bases 1 to 160716)  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
JOURNAL

**COMMENT**

Direct Submission  
 Submitted (08-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquery@sanger.ac.uk  
 On Jul 25, 2002 this sequence version replaced g1:21732123.  
 Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
 Project Information  
 Center project name: hb531C17  
 ----- Summary Statistics -----  
 Assembly program: XGAP4, version 4.5  
 Chemistry: Dye-terminator; 28 of reads  
 Chemistry: Dye-terminator; Big Dye; 97% of reads  
 Consensus quality: 158827 bases at least Q40  
 Consensus quality: 159096 bases at least Q30  
 Consensus quality: 159180 bases at least Q20  
 Insert size: 160716; sum-of-contigs  
 Insert size: 170984; 0.4% error; agarose-fp  
 Quality coverage: 10.84% in Q20 bases; sum-of-contigs  
 Quality coverage: 10.19% in Q20 bases; agarose-fp

## FEATURES

**Source**

misc\_feature

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP13-531c17"
/clone_id="RCIC-13.3"
1..166716
/contig="assembly-fragment:038771"
clone_end:566
vector_side:left
clone_end:77
vector_side:right

```

BASE COUNT	38997 a	41501 c	40997 g	39221 t
ORIGIN				

**Alignment Scores:**

1.44e-26

410.00

100.00%

100.008

22.928

2

US-09-748-451-2 (1-338) x AL691523 (1-160716)

QY 262 IleglulyslYsArgThrIleIleProThrLeuValGluAlaIleLysGluGlnAspGly 281

Db 144873 ATAGAAAGAACGCACCATCTCTACACTGGTGAAGCAATTAGGAACAAGATGGA 144932

QY 282 ArgGluValAspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLysLeu 301

Db 144933 AGAGAGTGGACTGGAGTATTTTATGGCCTGCTTTTACCTCAGAGAACCTAAACTA 144992

302 ValHisIleValCysHisLysLysThrThrHisLysLeuAsnCysAspProSerArgIle 321

Db 144993 GTGCATTTGTCTGCCATAGAAACCACCAAGCTCACTGTGACCCAGCAGATC 145052

QY 322 TYRLYSPROGLNTHRARGLEULYSARGLYSGLNPROVALARGLYSARGGLN 338

Db 145053 TACAACCCCGACAGGTTGAAGCGGAAGCAGCCTGTGCGGAACGCCAG 145103

RESULT 20

LOCUS	AK098413	1833 bp	mRNA	linear	PRI 15-JUL-2002
-------	----------	---------	------	--------	-----------------

ACCESSION AK098413  
VERSION 1 ST 01750410

**KEYWORDS** oligo capping; fls (full insert sequence).

clone: JTH01487.

Eukaryota; Metazoa; Chordata; Cranialta; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Placentalia; Carnivora; Canidae; Felidae

REFERENCE : 1  
ATTORNS Nino J. M. Costa v. Costa v. Vol-ol E

Kodaira, H., Furuya, T., Takahashi, M., Kikkawa, E., Omura, Y., Abe, K., Kato, Y., Katsuta, N., Goto, Y., Marikawa, M., Yamasaki, M.

Suzuki, Y., Hata, H., Nakagawa, K., Mizuno, S., Morinaga, M., Kawamura, M., Sugiyama, T., Ito, P., Otsuki, T., Sato, H.

Nishikawa, T., Sugiyama, A., Kawakami, B., Nagai, K., Isogai, T. and Sugano, S.

TITLE NEBO human cDNA sequencing project

REFERENCE 2 (bases 1 to 1833)  
ATTN: Drs. S. and Suzuki  
Suzuki, S. and Suzuki, Y.

**TITLE** Direct Submission  
**JOURNAL** Submitted (08-JUN-2002) Sumio Sugano Institute of Medical Science

Center: Shirokane-dai 4-6-1 Minato-ku Tokyo 108-8639 Japan

(E-mail: ccham1@ims.u-tokyo.ac.jp, Tel: 81-3-5449-3286,  
Fax: 81-3-5449-5416)

COMMENT: NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing:

Research Association for Biotechnology (RAB); cDNA library construction and 5'-end sequencing: Institute of Medical

Genome Center: 3'-end one pass sequencing: RAB: clone selection for science, University of Tokyo, Laboratory of Genome Structure, Human

FEATURES

Will insert sequencing: K&B and Helix Research Institute.

Location/Qualifiers

```
source
1: .1633
/oran1sm="Homo sapiens"
```

```
/usr_xie1=caxon:9606  
/c)one="JTH01487"
```

```
/cell_line="JCR"
/tissue_type="thyroid"
```



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2003, 15:21:08 ; Search time 121 Seconds  
(without alignments)  
277.028 Million cell updates/sec

Title: US-09-748-451-2  
Perfect score: 1789  
Sequence: 1 MLOKPSVKLRALRSPKRC.....SRITPQRLKQPVAKKQ 338

Gapop 10.0, Gapext 0.5  
375593 seqs, 9917265 residues

Total number of hits satisfying chosen parameters: 375593

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications, AA: \*  
1: /cgn2\_6/piddata1/pubpaa/US08\_NEM\_PUB.pep: \*  
2: /cgn2\_6/piddata1/pubpaa/PCIT\_NEM\_PUB.pep: \*  
3: /cgn2\_6/piddata1/pubpaa/US06\_NEM\_PUB.pep: \*  
4: /cgn2\_6/piddata1/pubpaa/US07\_NEM\_PUB.pep: \*  
5: /cgn2\_6/piddata1/pubpaa/US07\_NEM\_PUB.pep: \*  
6: /cgn2\_6/piddata1/pubpaa/US07\_PUBCOMB.pep: \*  
7: /cgn2\_6/piddata1/pubpaa/US08\_PUBCOMB.pep: \*  
8: /cgn2\_6/piddata1/pubpaa/US08\_PUBCOMB.pep: \*  
9: /cgn2\_6/piddata1/pubpaa/US08\_PUBCOMB.pep: \*  
10: /cgn2\_6/piddata1/pubpaa/US08\_PUBCOMB.pep: \*  
11: /cgn2\_6/piddata1/pubpaa/US10\_NEM\_PUB.pep: \*  
12: /cgn2\_6/piddata1/pubpaa/US10\_PUBCOMB.pep: \*  
13: /cgn2\_6/piddata1/pubpaa/US10\_PUBCOMB.pep: \*  
14: /cgn2\_6/piddata1/pubpaa/US10\_PUBCOMB.pep: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1789	100.0	338	US-09-748-451-2	Sequence 2, Appl
2	114	6.4	76	US-10-165-800-13	Sequence 13, Appl
3	99	5.5	466	US-10-242-043-12	Sequence 12, Appl
4	99	5.5	466	US-10-152-647-4	Sequence 4, Appl
5	97	5.4	405	US-10-043-487-302	Sequence 302, Appl
6	97	5.4	465	US-09-935-642-3	Sequence 3, Appl
7	97	5.4	466	US-09-923-779-155	Sequence 155, Appl
8	97	5.4	466	US-10-152-647-13	Sequence 13, Appl
9	91	5.1	219	US-10-165-800-14	Sequence 14, Appl
10	89	5.0	331	US-09-748-451-2	Sequence 2, Appl
11	89	5.0	457	US-10-086-135-3	Sequence 3, Appl
12	87	4.9	366	US-09-923-779-155	Sequence 155, Appl
13	86.5	4.8	219	US-10-165-800-11	Sequence 11, Appl
14	86	4.8	219	US-10-165-800-13	Sequence 13, Appl
15	85.5	4.8	1458	US-10-203-824-3	Sequence 3, Appl
16	84.5	4.7	373	US-10-143-849-2	Sequence 2, Appl
17	83	4.6	2120	US-10-051-311A-2	Sequence 2, Appl
18	82.5	4.6	277	US-09-867-550-1316	Sequence 1316, Appl
19	82.5	4.6	633	US-10-097-340-23	Sequence 23, Appl

20	82.5	4.6	718	US-10-097-340-19	Sequence 19, Appl
21	82.5	4.6	751	US-10-097-340-18	Sequence 18, Appl
22	81.5	4.6	628	US-10-151-193-6	Sequence 6, Appl
23	81.5	4.6	630	US-10-151-193-5	Sequence 5, Appl
24	81.5	4.6	725	US-10-118-328-5	Sequence 5, Appl
25	81	4.5	1457	US-09-772-316-2	Sequence 2, Appl
26	80	4.5	480	US-09-859-888-5	Sequence 5, Appl
27	78.5	4.4	1618	US-09-963-875-1	Sequence 1, Appl
28	78.5	4.4	1618	US-10-136-891-2	Sequence 1, Appl
29	78.5	4.4	1618	US-10-136-891-1	Sequence 1, Appl
30	78.5	4.4	1618	US-10-136-891-2	Sequence 1, Appl
31	78.5	4.4	1618	US-10-136-891-1	Sequence 1, Appl
32	77.5	4.3	1164	US-09-850-046A-2	Sequence 58, Appl
33	77.5	4.3	485	US-09-823-300-1664	Sequence 2, Appl
34	77.5	4.3	382	US-10-187-666A-48	Sequence 1664, Appl
35	77.5	4.3	382	US-09-987-107-37	Sequence 37, Appl
36	77.5	4.3	382	US-09-800-729-206	Sequence 206, Appl
37	77.5	4.3	629	US-10-197-666A-104	Sequence 104, Appl
38	77.5	4.3	629	US-09-799-875-17	Sequence 17, Appl
39	77.5	4.3	630	US-10-197-666A-102	Sequence 102, Appl
40	77.5	4.3	1331	US-09-801-368-370	Sequence 370, Appl
41	76.5	4.3	379	US-09-853-625B-11	Sequence 11, Appl
42	76.5	4.3	379	US-10-053-753-2	Sequence 2, Appl
43	76.5	4.3	679	US-10-278-547-42	Sequence 42, Appl
44	76.5	4.3	679	US-09-050-516-42	Sequence 42, Appl
45	76	4.2	1003	US-09-843-676-217	Sequence 217, Appl

ALIGNMENTS

US-09-748-451-2  
Sequence 2, Appl  
Patent No. US20010011078A1  
GENERAL INFORMATION:  
APPLICANT: WANG, XIAODONG  
APPLICANT: LIU, XUESONG  
TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN APOPTOSIS  
FILE REFERENCE: US20010011078A1  
CURRENT FILING DATE: 2000-12-22  
PRIORITY FILING DATE: 1998-04-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 338  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-748-451-2  
Query Match 100.0%; Score 1789; DB 10; Length 338;  
Best Local Similarity 100.0%; Pred. No. 8.3e-158;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MLOKPSVKLRALRSPKRCGVKRCGLRFPOLPERGSRICLYEGDTETDYF 60  
DB 1 MLOKPSVKLRALRSPKRCGVKRCGLRFPOLPERGSRICLYEGDTETDYF 60  
QY 61 PSYDPAELVLTICQAGQYSDIRPISAFHPRQVGLQAAQQLDDEAPQORILA 120  
DB 61 PSYDPAELVLTICQAGQYSDIRPISAFHPRQVGLQAAQQLDDEAPQORILA 120  
QY 121 DLHANTSONIAETRAEDPPFVEGLSFRSGSYLRYSCESTRINSYREVSPSYGA 180  
DB 121 DLHANTSONIAETRAEDPPFVEGLSFRSGSYLRYSCESTRINSYREVSPSYGA 180  
QY 181 EAOEELRVLYGSMQCRRLSMQYNSYEDRAGKGSRLCTPEGFGSCQGFMDSCLSRHS 240  
DB 181 EAOEELRVLYGSMQCRRLSMQYNSYEDRAGKGSRLCTPEGFGSCQGFMDSCLSRHS 240  
QY 241 INFSNRESRLTESWNLDTIEKKRTIPLTVEAIKEDQREVMEXFYGLFTSENK 300

Db 241 INPISNRSLRILFTSWIMDLHIEKKRITLPLVPAIKQDREVDWETFGILFTSEMLK 300  
QY 301 LVHIVCHKKTKHLNCDPSRIYKQPTRLKRPYARRKQ 338  
Db 301 LVHIVCHKKTKHLNCDPSRIYKQPTRLKRPYARRKQ 338

## RESULT 2

US-10-165-800-13  
Sequence 13, Application US/10165800  
Publication No. US20030092116A1  
GENERAL INFORMATION:  
APPLICANT: Chun, Miyoung  
APPLICANT: Glucksmann, Maria Alexandra  
APPLICANT: Kapeller-Libermann, Rosana  
APPLICANT: Meyers, Rachel E.  
TITLE OF INVENTION: No. US20030092116A1 Nucleic Acid Sequences Encoding  
TITLE OF INVENTION: Adenylate kinase, Phospholipid Scramblase-Like, DNA  
TITLE OF INVENTION: Fragmentation Factor-Like, Phosphatidylserine Synthase-Like,  
TITLE OF INVENTION: and ATPase-Like Molecules and Uses Therefor  
FILE REFERENCE: 35800/247838  
CURRENT APPLICATION NUMBER: US/10/165,800  
CURRENT FILING DATE: 2002-06-07

PRIOR APPLICATION NUMBER: 09/790,179  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/185,609  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: 09/790,838  
PRIOR FILING DATE: 2001-02-22  
PRIOR APPLICATION NUMBER: 60/185,946  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: 09/790,180  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/185,947  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: 09/795,038  
PRIOR FILING DATE: 2001-02-26  
PRIOR APPLICATION NUMBER: 60/186,234  
PRIOR FILING DATE: 2000-02-29  
PRIOR APPLICATION NUMBER: 09/781,677  
PRIOR FILING DATE: 2001-02-21  
PRIOR APPLICATION NUMBER: 60/181,705  
PRIOR FILING DATE: 2000-02-10  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 76  
TYPE: PRT  
ORGANISM: Artificial Sequence

FEATURE:  
OTHER INFORMATION: Consensus sequence for the CAD domain  
-165-800-13

Query Match  
Best Local Similarity 6.4%; Score 114; DB 9; Length 76;  
Matches 30; Conservative 12; Mismatches 22; Indels 4; Gaps 3;

QY 14 RSPRRKFGVAGRSCEVLKRGKLPQPE--RGSRLCLYEDGTETL--TEDYPSVDNAELV 70  
Db 10 RNVAK--GVAASSTLELLSKVLKLPDLSLEPVTLVLEEDGTETVDEDEYFTLNNELV 68  
QY 71 LRLTGAW 78  
Db 69 ALDEGKRW 76

## RESULT 3

US-10-242-943-12  
Sequence 12, Application US/10242943  
Publication No. US20030087412A1  
GENERAL INFORMATION:  
APPLICANT: Nandabalan, Krishnan

APPLICANT: Schulz, Vincent P.  
APPLICANT: Yang, Melja  
TITLE OF INVENTION: NIK1 PROTEIN AND NIK1 PROTEIN COMPLEXES  
FILE REFERENCE: 15966-521 NIK1 protein complexes  
CURRENT APPLICATION NUMBER: US/10/242,943  
CURRENT FILING DATE: 2002-09-13  
PRIOR APPLICATION NUMBER: US/09/167,206  
PRIOR FILING DATE: 1998-10-06  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 12  
LENGTH: 466  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-242-943-12

Query Match  
Best Local Similarity 5.5%; Score 99; DB 9; Length 466;  
Matches 65; Conservative 47; Mismatches 111; Indels 90; Gaps 15;

QY 27 QEVLRKGLRQLPERSRLCLYEDGTETLTDYPSVDNAELVLTGQAWGVSADIR 86  
Db 180 EDIMR---LRKLGEE---MLGREAEANTLQSFROVDNNSLARLDLKRVESLQGEI 231  
QY 87 RFLSAFHPPQVGLQQA-----QQLCDQAPQORRLADLLHNVSONIAETRAEDPPW 141  
Db 232 AFLKRLHEETIQELQAOIQEQHVQIDVDSKPDITLALRDVROQY--ESVAAKNLQAEWE 290  
QY 142 F-----EGLESRFQSKSGLYR-----SCSRIR-----S 166  
Db 291 YKSFADLSEANNNALRKAKOESTETFRQVOSLCEVDALKTGESLERORKEEN 350  
QY 167 YLREVSSTPVGAQAEFLVLSMCQRLRSQVNGSYFDGAKGSGRLCTPEGWFS 226  
Db 351 FAVEAANYQDTIG--RLQDEIQNMKEAMRHLREYQ-----DLLNV 389  
QY 227 QGPPDMSCLSRHSINPISNRSLKIL-----FST-----WNLDI-----IEKRTIIP 272  
Db 390 KMALDIEIATYRKIL---EGESRISLPLPNTSSILNRETMLDLSLPVDTSHSKRTL---L 443  
QY 273 VEAIKEDDGRVD 285  
Db 444 IKIVETRDGVIN 456

## RESULT 4

US-10-152-647-4  
Sequence 4, Application US/10152647  
Patent No. US20020137110A1  
GENERAL INFORMATION:  
APPLICANT: MORISHIMA, No. US20020137110A1uhhro,  
APPLICANT: NAKANISHI, Keiko,  
APPLICANT: SHIBATA, Takehiko

TITLE OF INVENTION: Antibody against cleavage product of vimentin  
FILE REFERENCE: 522.1004  
CURRENT APPLICATION NUMBER: US/10/152,647  
CURRENT FILING DATE: 2002-05-21  
PRIOR APPLICATION NUMBER: US/09/610,401  
PRIOR FILING DATE: 2000-07-05  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 466  
TYPE: PRT  
ORGANISM: Mus sp.  
US-10-152-647-4

Query Match  
Best Local Similarity 5.5%; Score 99; DB 12; Length 466;  
Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

QY 27 QEVLRKGLRQLPERSRLCLYEDGTETLTDYPSVDNAELVLTGQAWGVSADIR 86  
Db 27 QEVLRKGLRQLPERSRLCLYEDGTETLTDYPSVDNAELVLTGQAWGVSADIR 86

Db 180 EDIMR--LREKIOAE---MLQREAEESTLQSFROVDNANSLARLDLERRKVESLOEET- 231  
QY 87 RLTSAFHEPOVGLIOA-----OQLLCDEQAPORRLADLLHNVSQNIATETRAEDPPW 141  
Db 233 AFLKRLHEEIOELQAOIOEOHQVIDVSKPDLTAALRDVROQY-ESVAAKNIOEAEEM 290  
QY 142 F-----EGLESRFQSKGYLRY-----SCESRIR-----S 166  
Db 291 YKSKFADLSEANNNNDLROAKQESNEYRQVQSLTCEVDALKTNLSEROMREMEEN 350  
QY 167 YLREVSYSYTGAEAOEFLRYLGSRCORLRSMQYNGSYFDRGAKGSRILCTPEGWESC 226  
Db 351 FALLEANQDITIG-RLODEIQNMKEEMARHLREYO-----DLNV 389  
QY 227 QGPFMDSCLSRHSINPYSNRESRL-----LFTWNLDHI-----IEKRTIIPTL 272  
Db 390 KMAIDIEIATYRKL-----EGESRISLPLPFSLSLNETNLSLPLVDTHSKRTL---L 443  
QY 273 VEAIKRODQREVD 285  
Db 444 IKVETRDQVYN 456

RESULT 5  
US-10-043-487-302  
; Sequence 302, Application US/10043487  
; Publication No. US20030055220A1  
; GENERAL INFORMATION:  
; APPLICANT: HYBRIGENICS  
; APPLICANT: Pierre, Legrain  
; TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptide  
; FILE REFERENCE: B4778A  
; CURRENT APPLICATION NUMBER: US/10/043,487  
; PRIOR FILING DATE: 2002-04-30  
; PRIOR APPLICATION NUMBER: US 60/261,130  
; PRIOR FILING DATE: 2001-01-12  
; NUMBER OF SEQ ID NOS: 561  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 302  
; LENGTH: 405  
; TYPE: PRT  
; ORGANISM: Shigella flexneri  
US-10-043-487-302

Query Match 5.4%; Score 97; DB 9; Length 405;  
Best Local Similarity 20.8%; Pred. No. 0.87; Indels 90; Gaps 15;  
Matches 65; Conservative 46; Mismatches 112;  
Db 27 QEVLKRGCLRFQLPERSRLCLYEDGTETEDYFSPVPMNAELVLTIGQAMQGYSDIR 86  
QY 119 EDIMR--LREKIOAE---MLQREAEENTLQSFROVDNANSLARLDLERRKVESLOEET- 170  
Db 87 RLTSAFHEPOVGLIOA-----OQLLCDEQAPORRLADLLHNVSQNIATETRAEDPPW 141  
QY 171 AFLKRLHEEIOELQAOIOEOHQVIDVSKPDLTAALRDVROQY-ESVAAKNIOEAEEM 229  
Db 142 F-----EGLESRFQSKGYLRY-----SCESRIR-----S 166  
QY 230 YKSKFADLSEANNNNDLROAKQESNEYRQVQSLTCEVDALKTNLSEROMREMEEN 289  
QY 167 YLREVSYSYTGAEAOEFLRYLGSRCORLRSMQYNGSYFDRGAKGSRILCTPEGWESC 226  
Db 290 FAVEAANYQDITIG-RLODEIQNMKEEMARHLREYO-----DLNV 328  
QY 227 QGPFMDSCLSRHSINPYSNRESRL-----LFTWNLDHI-----IEKRTIIPTL 272  
Db 329 KMAIDIEIATYRKL-----EGESRISLPLPFSLSLNETNLSLPLVDTHSKRTL---L 382  
QY 273 VEAIKRODQREVD 285  
Db 383 IKVETRDQVYN 395

RESULT 6  
US-09-935-642-3  
; Sequence 3, Application US/09935642  
; Publication No. US20030044795A1  
; GENERAL INFORMATION:  
; APPLICANT: BYRJAENSEN, Inger  
; APPLICANT: LARSEN, Peter  
; APPLICANT: STEPHEN, John  
; TITLE OF INVENTION: Biochemical Markers for the Human  
; FILE REFERENCE: 8969-014  
; CURRENT APPLICATION NUMBER: US/09/935,642  
; PRIOR FILING DATE: 2001-08-24  
; CURRENT APPLICATION NUMBER: PCT/GB97/02394  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: PCT/GB97/07132.8  
; PRIOR FILING DATE: 1997-04-08  
; PRIOR APPLICATION NUMBER: PCT/GB9618600.2  
; PRIOR FILING DATE: 1996-09-06  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-935-642-3

Query Match 5.4%; Score 97; DB 9; Length 465;  
Best Local Similarity 20.8%; Pred. No. 1.1; Indels 90; Gaps 15;  
Matches 65; Conservative 46; Mismatches 112;  
Db 27 QEVLKRGCLRFQLPERSRLCLYEDGTETEDYFSPVPMNAELVLTIGQAMQGYSDIR 86  
QY 179 EDIMR--LREKIOAE---MLQREAEENTLQSFROVDNANSLARLDLERRKVESLOEET- 230  
Db 87 RLTSAFHEPOVGLIOA-----OQLLCDEQAPORRLADLLHNVSQNIATETRAEDPPW 141  
QY 231 AFLKRLHEEIOELQAOIOEOHQVIDVSKPDLTAALRDVROQY-ESVAAKNIOEAEEM 289  
Db 142 F-----EGLESRFQSKGYLRY-----SCESRIR-----S 166  
QY 290 YKSKFADLSEANNNNDLROAKQESNEYRQVQSLTCEVDALKTNLSEROMREMEEN 349  
QY 167 YLREVSYSYTGAEAOEFLRYLGSRCORLRSMQYNGSYFDRGAKGSRILCTPEGWESC 226  
Db 350 FAVEAANYQDITIG-RLODEIQNMKEEMARHLREYO-----DLNV 388  
QY 227 QGPFMDSCLSRHSINPYSNRESRL-----LFTWNLDHI-----IEKRTIIPTL 272  
Db 389 KMAIDIEIATYRKL-----EGESRISLPLPFSLSLNETNLSLPLVDTHSKRTL---L 442  
QY 273 VEAIKRODQREVD 285  
Db 443 IKVETRDQVYN 455

RESULT 7  
US-09-923-779-155  
; Sequence 155, Application US/09923779  
; Patent No. US20020076721A1  
; GENERAL INFORMATION:  
; APPLICANT: Pyle, Ruth A.  
; APPLICANT: Xu, Jiangchun  
; APPLICANT: Kelos, Michael D.  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
; FILE REFERENCE: 210121.553  
; CURRENT APPLICATION NUMBER: US/09/923,779  
; PRIOR FILING DATE: 2001-08-06  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 155

LENGTH: 466  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-923-779-155

Query Match 5.48; Score 97; DB 10; Length 466;  
Best Local Similarity 20.88; Pred. No. 1.1;  
Matches 65; Conservative 46; Mismatches 112; Indels 90; Gaps 15;

QY 27 QEVLRKGLRFPOLPERGSRCLYEDGTELEDYFPPVDPNAELVLTGQAMQGYSDIR 86  
DB 180 EDIMR---LREKIQEE---MLQREAEENTLQSFQVDVNASLARLDLEKRVESIQEEI- 231  
QY 87 RFLSAFHEPQVGLIOA-----OQLCDEQAPQORLLADLHNVSQNIATRAEDPPW 141  
DB 232 AFLKKEHEEIOELQIOEQHVQIDVDVSKPDLTALRDVROQY-ESVAKNQLQEAEMW 290  
QY 142 F-----EGLESRFQSKSGYLRY-----SCESRIR-----S 166  
DB 291 YKSKFADLSEANRNDALROAKQOESTERYRQVOSLTCEVDALKGTNESLEROMREEM 350  
QY 167 YLREVSSTPYGVAEAOEFLRVLGSQCRLRSQVNGSYFDRGAKGSRCLTPEGWFS 226  
DB 351 FAVEAANYODTIG-RLODEIQNMKEEMARHLREYQ-----DLLNV 389  
QY 227 QGFPDMDCSLSRHSINYSRRESRIL-----FST-----WNLDHI-----TEKRTIIP 272  
DB 390 KMALDIEIATYRKLL---EGEESRISLPLPFSLSLNRETNLDLSLPLVDTHSKRTF---L 443  
QY 273 VEAKEODGREVD 285  
DB 444 IKTVETRDGOVIN 456

RESULT 8  
US-10-152-647-3

; Sequence 3, Application US/10152647  
; Patent No. US20020137110A1  
; GENERAL INFORMATION:  
; APPLICANT: MORISHIMA, No. US20020137110A1uh1ro,  
; APPLICANT: NAKANISHI, Keiko,  
; APPLICANT: SHIBATA, Takehiko  
; TITLE OF INVENTION: Antibody against cleavage product of vimentin  
; FILE REFERENCE: 522.1004  
; CURRENT APPLICATION NUMBER: US/10/152,647  
; CURRENT FILING DATE: 2002-05-21  
; PRIOR APPLICATION NUMBER: US/09/610,401  
; PRIOR FILING DATE: 2000-07-05  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 3  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-152-647-3

Query Match 5.48; Score 97; DB 12; Length 466;  
Best Local Similarity 20.88; Pred. No. 1.1;  
Matches 65; Conservative 46; Mismatches 112; Indels 90; Gaps 15;

QY 27 QEVLRKGLRFPOLPERGSRCLYEDGTELEDYFPPVDPNAELVLTGQAMQGYSDIR 86  
DB 180 EDIMR---LREKIQEE---MLQREAEENTLQSFQVDVNASLARLDLEKRVESIQEEI- 231  
QY 87 RFLSAFHEPQVGLIOA-----OQLCDEQAPQORLLADLHNVSQNIATRAEDPPW 141  
DB 232 AFLKKEHEEIOELQIOEQHVQIDVDVSKPDLTALRDVROQY-ESVAKNQLQEAEMW 290  
QY 142 F-----EGLESRFQSKSGYLRY-----SCESRIR-----S 166  
DB 291 YKSKFADLSEANRNDALROAKQOESTERYRQVOSLTCEVDALKGTNESLEROMREEM 350  
QY 167 YLREVSSTPYGVAEAOEFLRVLGSQCRLRSQVNGSYFDRGAKGSRCLTPEGWFS 226

DB 351 FAVEAANYODTIG-RLODEIQNMKEEMARHLREYQ-----DLLNV 389  
QY 227 QGFPDMDCSLSRHSINYSRRESRIL-----FST-----WNLDHI-----TEKRTIIP 272  
DB 390 KMALDIEIATYRKLL---EGEESRISLPLPFSLSLNRETNLDLSLPLVDTHSKRTF---L 443  
QY 273 VEAKEODGREVD 285  
DB 444 IKTVETRDGOVIN 456

RESULT 9  
US-10-165-800-14

; Sequence 14, Application US/10165800  
; Publication No. US20030092116A1  
; GENERAL INFORMATION:  
; APPLICANT: Chun, Myoung  
; APPLICANT: Glucksmann, Maria Alexandra  
; APPLICANT: Kapeller-Libermann, Rosana  
; APPLICANT: Meyers, Rachel E.  
; TITLE OF INVENTION: No. US20030092116A1el Nucleic Acid Sequences Encoding  
; TITLE OF INVENTION: Adenylate Kinase, Phospholipid scramblase-Like, DNA  
; TITLE OF INVENTION: Fragmentation Factor-Like, Phosphatidylserine Synthase-Like,  
; FILE REFERENCE: 35800/247838  
; CURRENT APPLICATION NUMBER: US/10/165,800  
; PRIOR FILING DATE: 2002-06-07  
; PRIOR APPLICATION NUMBER: 09/790,179  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/185,609  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: 09/790,838  
; PRIOR FILING DATE: 2001-02-22  
; PRIOR APPLICATION NUMBER: 60/185,946  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: 09/790,180  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/185,947  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: 09/795,038  
; PRIOR FILING DATE: 2001-02-26  
; PRIOR APPLICATION NUMBER: 60/186,234  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: 09/781,677  
; PRIOR FILING DATE: 2001-02-21  
; PRIOR APPLICATION NUMBER: 60/181,705  
; PRIOR FILING DATE: 2000-02-10  
; NUMBER OF SEQ ID NOS: 25  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 14  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-10-165-800-14

Query Match 5.18; Score 91; DB 9; Length 219;  
Best Local Similarity 25.88; Pred. No. 1.4;  
Matches 46; Conservative 21; Mismatches 71; Indels 40; Gaps 7;

QY 14 RSPRKFVAVRSQOEVLARKGLRFPOLPERGSRCLYEDGTELEDYFPPVDPNAELVLT 72  
DB 45 RYARK-GLTAAISOELDK-VLETLRLRGVLTVLEEDGYAVDSEDFQILLEDDTCIMVL 102  
QY 73 TLQAW-----OGYVS-----DIRFLSAFHEPQ-----VGLIOAQQIL 107  
DB 103 EOGQSWPKSGMLSTYGGRKPKHSDIAITTDVYKQNRDLFGSLNVAATYTGYSMS 162  
QY 108 CDEQAPQORLLADLHNVSQNIATRAEDPPWEGLESRFQSKSGYLRYSCESRIR 165  
DB 163 CDQGVGPKRVYLLBELLNGTSSQL-----QGLGMLLGISSTLKHVYVAGADR 208

RESULT 10  
US-09-748-451-4  
Sequence 4, Application US/09748451  
Patent No. US20010011078A1  
GENERAL INFORMATION:  
APPLICANT: WANG, XINQONG  
APPLICANT: LIU, XUESONG  
TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN APOPTOSIS  
FILE REFERENCE: UTSD:5460SD1  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/061,702  
PRIOR FILING DATE: 1998-04-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 4  
LENGTH: 331  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-748-451-4

Query Match 5.0%; Score 89; DB 10; Length 331;  
Best Local Similarity 22.2%; Pred. No. 3.7; Mismatches 146; Indels 78; Gaps 15;  
Matches 78; Conservative 49; Mismatches 146; Indels 78; Gaps 15;  
QY 5 PKSVKRLAIR-----SPKRGVAGRSCEVLR-KGC--LRFQLPERSRLCYEDGT 53  
DB 10 PESGEIRTLKPCLLRRNYSRQHGVA--SLEDLRKACDILAIDKSLTPVTLVAEDGT 68  
QY 54 EL-TEDYFSPVDPNAELVLTGQAMQGVSDIRRLSAPHEPOVGLIOAAQQLCDEQA 112  
DB 69 IYDDDYFLCLPSNKKFVALASNEKW-----AYNSDGTAMISQESFDVET 116  
QY 113 PORRLADLHNVSONIAETRAEDPPWFGLESFRQSGYLRSCTRSRLRYREVS 172  
DB 117 DSGAGL-----KKKNVARELK-----EDLSSTLLSEEDLQMLVDACSDLADEL 162  
QY 173 SYPTGAEAGEEFLVIGSMQRLRSQYNSYFDRGAKGSRCLCTEGWFGSCGPPD- 231  
DB 163 OSCATV-ORLQHTLQOVLDOREEVRSQKLLQLYLQALKEGSLSKGE--ESKAARGE 218  
QY 232 ---MDSCLSRHSINPYNSRESRLFTSNLHDHIIKKRTIIPVLEAIKDDGREVDWE 287  
DB 219 EYDAVDGTMSRET-----SSDVALLS-----HIL-----TALREKQAE----- 252  
QY 288 YFGGLTFSENKIVHYCHKKTKTHKNCDSRLTKRQPRKQ 338  
DB 253 ---LSLSQDLLEVTKEDPFAVALNMDIKKTEYVQACERELALRLQ 299

RESULT 11  
US-10-086-135-3  
Sequence 3, Application US/10086135  
Publication No. US20020182677A1  
GENERAL INFORMATION:  
APPLICANT: Sheppard, Paul O.  
APPLICANT: Chandrasekhar, Yashin A.  
TITLE OF INVENTION: PANCREATIC AND OVARIAN POLYPEPTIDE  
FILE REFERENCE: ZS1G58  
CURRENT FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: US/09/086,135  
PRIOR FILING DATE: 2002-02-26  
PRIOR APPLICATION NUMBER: US/09/366,448  
PRIOR FILING DATE: EARLIER FILING DATE: 1999-08-03  
PRIOR APPLICATION NUMBER: EARLIER FILING DATE: 1998-06-03  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 3  
LENGTH: 457  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-10-086-135-3

Query Match 5.0%; Score 89; DB 9; Length 457;  
Best Local Similarity 21.8%; Pred. No. 5.7; Mismatches 104; Gaps 16;  
Matches 60; Conservative 40; Mismatches 104; Indels 104; Gaps 16;  
QY 44 SRLCYEDGTETEDYFSPVDPNAELVLTGQAMQGVSDIRRLSAPHEPOV--VGLI 100  
DB 9 SLVTLVWGTGLTQ-----VLPNPE-----ESWOY-----SSAQDSGRCITV 49  
QY 101 QAAQQLCDEQAPORO-BLLADLHNVSONIAETRA--AEDPPWFE-----GLESR- 149  
DB 50 VAPQGMCSRDATKQRLKQVNMOSIEYLDRTQRLDQYKEMQKGLSEKFR 109  
QY 150 -----OSKSGYL-----YSCSR-IRSYLREVSYPST----- 177  
DB 110 QVEESHQHLARQKAKRKMDELRLPILVLEEKADAKIVLOFKEEVQNLTSVNLQ 169  
QY 178 -VGAEOEFLVIGSMQRLR-----SMQYNSYFDRGAKGSR- 217  
DB 170 EIGAVDYDELQSRVSNLEERLQACQKLAGCKLGLSDPYVTWTSGR-----GSMNT 223  
QY 218 --CPRG---WFGCGPPDMDSCLSRHSINPYNS 246  
DB 224 DPLAEPEDNRVY-----MDGYHNRFTVEYKNS 251

RESULT 12  
US-09-925-300-1495  
Sequence 1495, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Ruben,  
APPLICANT: Steve Ruben,  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1495  
LENGTH: 366  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-300-1495

Query Match 4.9%; Score 87; DB 10; Length 366;  
Best Local Similarity 22.4%; Pred. No. 6.4; Mismatches 131; Indels 100; Gaps 18;  
Matches 81; Conservative 50; Mismatches 131; Indels 100; Gaps 18;  
QY 5 PKSVKRLAIR-----SPKRGVAGRSCEVLR-KGC--LRFQLPERSRLCYEDGT 53  
DB 45 PESGEIRTLKPCLLRRNYSRQHGVA--SLEDLRKACDILAIDKSLTPVTLVAEDGT 103  
QY 54 EL-TEDYFSPVDPNAELVLTGQAMQGVSDIRRLSAPHEPOVGLIOAAQQLC-----C 108  
DB 104 IYDDDYFLCLPSNKKFVALASNEKW-----AYNSDGTAMISQESFDVET 151  
QY 109 DEQAPORRLADLHNVSONIAETRAE-----DPPWFGLSRQSGYLRSCTRS- 162  
DB 152 DSGAGLKKWKNVAKQLKEDLSITLLSEEDLQMLVDACSDLAQ-----LRQCATV 203  
QY 163 -RISYLRREVSYPVGAEOEFLVIGSMQRLRSQYNSYFDRGAKGSRCLCTE 221  
DB 204 QRLQHTLQOVL-----QREVR-----QSKLLDLYLQALKEGSLSKGE 245  
QY 222 GWFSCGPPD-----MDSCLSRHSINPYNSRESRLFTSNLHDHIIKKRTIIPVLA 276  
DB 246 ---ESKAARGEYDAVDGTISRET-----SSDVALLS-----HIL-----TAL 280

QY 277 KEDGKREVDWETFFGLFTSENKLVAVICHTTKTKINCPSRIYKQTLKKRPYK 336  
Db 281 REKQAP-----LSLSODELVKEDPKALAVALNMDIKRTETVQACERELARL 332  
QY 337 RQ 338  
Db 333 QQ 334

## RESULT 13

US-10-165-800-11  
; Sequence 11, Application US/10165800  
; Publication No. US20030092116A1

## GENERAL INFORMATION:

APPLICANT: Chun, Miyoung  
APPLICANT: Glucksmann, Maria Alexandra  
APPLICANT: Kapeller-Liebermann, Rosana  
APPLICANT: Meyers, Rachel E.  
TITLE OF INVENTION: No. US20030092116A1el Nucleic Acid Sequences Encoding  
TITLE OF INVENTION: Adenylate kinase, Phospholipid Scramblase-Like, DNA  
TITLE OF INVENTION: Fragmentation Factor-Like, and Phosphatidylserine Synthase-Like,  
FILE REFERENCE: 35800/247838  
CURRENT APPLICATION NUMBER: US/10/165,800

;; PRIOR FILING DATE: 2002-06-07  
;; PRIOR APPLICATION NUMBER: 09/790,179  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/185,609  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/790,838  
;; PRIOR FILING DATE: 2001-02-22  
;; PRIOR APPLICATION NUMBER: 60/185,946  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/790,180  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/185,947  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/795,038  
;; PRIOR FILING DATE: 2001-02-26  
;; PRIOR APPLICATION NUMBER: 60/186,234  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/781,677  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/181,705  
;; PRIOR FILING DATE: 2000-02-10  
;; NUMBER OF SEQ ID NOS: 25  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 11  
;; LENGTH: 219  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; 10-165-800-11

Query Match 4.8%; Score 86.5; DB 9; Length 219;  
Best Local Similarity 24.9%; Pred. No. 3.5;  
Matches 47; Conservative 25; Mismatches 68; Indels 49; Gaps 8;

QY 14 RSPKFGVAGRSCEVLKGLRQLPERSRLCLYEDGTEL-TEDYFSPVDNAELVYL 72  
Db 45 RTIRK-GILTAATROELAK-ALETLLNGVLTVLVEEDGTAVDSEDFQLEDDTCLAVL 102  
QY 73 TLGAW-----QGYVS-----DIRFLSAFHEPQ-----VGLIQAAQOL 106  
Db 103 QSGQSWSPTRSGVSLGIGRERPRKSRDIAFTFDVYKQNRDLFGSLNVAATPYGLYSM 162  
QY 107 LCDQAOAORORLLDLHNVSQNTAETRADPPMEGESLESFQSGEYLRYSGE----- 161  
Db 163 SCDDQGGGPKKVLVELLRLWTSTL-----QGLGMILGISTLTHAAGAQQW 210  
QY 162 ---SRISY 167  
Db 211 QOKGRHSY 219

## RESULT 14

US-10-165-800-15  
; Sequence 15, Application US/10165800  
; Publication No. US20030092116A1

## GENERAL INFORMATION:

APPLICANT: Chun, Miyoung  
APPLICANT: Glucksmann, Maria Alexandra  
APPLICANT: Kapeller-Liebermann, Rosana  
APPLICANT: Meyers, Rachel E.  
TITLE OF INVENTION: No. US20030092116A1el Nucleic Acid Sequences Encoding  
TITLE OF INVENTION: Adenylate kinase, Phospholipid Scramblase-Like, DNA  
TITLE OF INVENTION: Fragmentation Factor-Like, and Phosphatidylserine Synthase-Like,  
FILE REFERENCE: 35800/247838  
CURRENT APPLICATION NUMBER: US/10/165,800

;; PRIOR FILING DATE: 2002-06-07  
;; PRIOR APPLICATION NUMBER: 09/790,179  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/185,609  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/790,838  
;; PRIOR FILING DATE: 2001-02-22  
;; PRIOR APPLICATION NUMBER: 60/185,946  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/790,180  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/185,947  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/795,038  
;; PRIOR FILING DATE: 2001-02-26  
;; PRIOR APPLICATION NUMBER: 60/186,234  
;; PRIOR FILING DATE: 2000-02-29  
;; PRIOR APPLICATION NUMBER: 09/781,677  
;; PRIOR FILING DATE: 2001-02-21  
;; PRIOR APPLICATION NUMBER: 60/181,705  
;; PRIOR FILING DATE: 2000-02-10  
;; NUMBER OF SEQ ID NOS: 25  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 15  
;; LENGTH: 219  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; US-10-165-800-15

Query Match 4.8%; Score 86; DB 9; Length 219;  
Best Local Similarity 36.4%; Pred. No. 3.9;  
Matches 24; Conservative 11; Mismatches 29; Indels 2; Gaps 2;

QY 14 RSPKFGVAGRSCEVLKGLRQLPERSRLCLYEDGTEL-TEDYFSPVDNAELVYL 72  
Db 44 RSSRR-GVMASLSQELISKTDALVATGLVTLVEEDGTVVDTEEFQTLGDMTHFWIL 102  
QY 73 TLGAW 78  
Db 103 EKGQKW 108

RESULT 15  
US-10-203-224-3  
; Sequence 3, Application US/10203224  
; Publication No. US20030086945A1

GENERAL INFORMATION:  
APPLICANT: FAABERG, James E.  
APPLICANT: ROSSOW, Kurt D.  
TITLE OF INVENTION: PROLINE REPRODUCTIVE AND RESPIRATORY SYNDROME VIRUS AND  
FILE REFERENCE: 110,01250101  
CURRENT APPLICATION NUMBER: US/10/203,224  
CURRENT FILING DATE: 2002-08-07  
PRIOR APPLICATION NUMBER: PCY/US01/04351  
PRIOR FILING DATE: 2001-02-08

PRIOR APPLICATION NUMBER: 60/181,041  
PRIOR FILING DATE: 2000-02-08  
PRIOR APPLICATION NUMBER: 60/193,220  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: 60/206,624  
PRIOR FILING DATE: 2000-05-24  
PRIOR APPLICATION NUMBER: 60/215,373  
PRIOR FILING DATE: 2000-06-29  
PRIOR APPLICATION NUMBER: 60/260,041  
PRIOR FILING DATE: 2001-01-05  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO: 3  
LENGTH: 1458  
TYPE: PRT  
ORGANISM: Porcine reproductive and respiratory syndrome virus  
0-203-224-3

Query Match 4.8%; Score 85.5; DB 9; Length 1458;  
Best Local Similarity 22.1%; Pred. No. 59;  
Matches 64; Conservative 36; Mismatches 96; Indels 93; Gaps 15;

QY 100 IQAQQQLCDEQAPORQLADLHNSQNTAETRA-----BDPWFEGCL 145  
DB 553 IEAGRLV-----PNDRLALATYHMAQNASETYASAAIIMDSACIDHDPEMTDL 607  
QY 146 ---ESRFQSKGY-----LRYSCSRIRSYLREVSYPSTVGAQAQEEFLRYLG-SKQ 195  
DB 608 ICGIAQACQADGYSFPPAFEMSWMERLRSH-NEGKKFRHCGICDAKADYASACGLDLC- 665  
QY 196 RLRSQYNGSYFDR-----GANGSRILCTPGWFSGCGPF-----DMSCLSRHSIN 242  
DB 666 -----LFHSHFHQHPVYLSCGHAGSRECS-----QCOSPYGAGRSPLDVLQOIPYK 714  
QY 243 PYSNRESRIILFTWNLD-----HIEKKRTIIPLVLAIREQDGRVDE---YFY 290  
DB 715 PPRVIVKVGKNTTALDPGRYQSRRLVAVKRGIA-----GNEVDSIDGQVY 763

QY 291 GLFPTSENLKLVHVGCH-----KKTTHKIN--CDPSRIYKP 324  
DB 764 PLLPCTCKDIMWAKVACNVLLSKFTYGPFGSKTTWLLGQVODDVIYTP 812

RESULT 16  
US-10-143-849-2  
Sequence 2, Application US/10143849  
Publication No. US20030074699A1  
GENERAL INFORMATION:  
APPLICANT: Coupland, George M.  
Putterill, Joanna J.  
TITLE OF INVENTION: Genetic control of flowering  
NUMBER OF SEQUENCES: 31  
CORRESPONDENCE ADDRESS:  
ADDRESS: Nixon & Vanderhye PC  
STREET: 8th Floor, 1100 No. US20030074699A1th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: United States of America  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/143,849  
FILING DATE: 14-May-2002  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/09/516,191  
FILING DATE: <unknown>  
APPLICATION NUMBER: 08/945,056  
FILING DATE: <unknown>

APPLICATION NUMBER: GB 9422083.7  
FILING DATE: 02-NOV-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: MS MARY J WILLSON  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-17  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 373 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-10-143-849-2

Query Match 4.7%; Score 84.5; DB 9; Length 373;  
Best Local Similarity 22.6%; Pred. No. 11;  
Matches 47; Conservative 26; Mismatches 64; Indels 71; Gaps 10;

QY 153 SGYLRSCSRIRSYLREVSYPSTVGAQAQEEFLRYLGMCRLSMQYNGSYFDRGAK 212  
DB 36 SAYLCMSQDAQVHSANRVAS-----RHKRVYCES-CERAPV----- 71  
QY 213 GGSRLCTPGWFSGCGPFDMDSCLSR-HSINPYSNRESRI-----LESTWNLHIE 263  
DB 72 -ATLCADADASIC-----TACDEVSANPLARRHQVPLIPISGNSFSMTTHQS 123  
QY 264 KKRTIIPVLAIREQDGRVDE---WEY-----FYGLFSTENKIV--- 302  
DB 124 EKMTDEKRLVLDQDESGEDGDAKEVASLFLPNSKNNNNNGLLFSDRYNLVDYN 183  
QY 303 ---HIVCAKKTTHKINCDPSRIYKP 326  
DB 184 SMDKFTGETGYEQHQDNC-----VPQT 206

RESULT 17  
US-10-051-311A-2  
Sequence 2, Application US/10051311A  
Patent No. US20020164620A1  
GENERAL INFORMATION:  
APPLICANT: Boehringer Ingelheim International GmbH  
TITLE OF INVENTION: Method for identifying compounds that modulate sister  
FILE REFERENCE: 0652, 2290001  
CURRENT APPLICATION NUMBER: US/10/051,311A  
PRIOR FILING DATE: 2002-06-20  
PRIOR APPLICATION NUMBER: EP 01 101 252.3  
PRIOR FILING DATE: 2001-01-19  
PRIOR APPLICATION NUMBER: US 60/297,440  
NUMBER OF SEQ ID NOS: 13  
SOFTWARE: Patent Ver. 2.1  
SEQ ID NO: 2  
LENGTH: 2120  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-051-311A-2

Query Match 4.6%; Score 83; DB 9; Length 2120;  
Best Local Similarity 18.8%; Pred. No. 1.7e+02;  
Matches 66; Conservative 52; Mismatches 130; Indels 104; Gaps 14;

QY 11 RALRSRPFAGVAGNSCOE-----VLRKGLRFPQLPENGSS-----RLCYEDGT 53  
DB 197 RAVAAHQFLDASGHGNEADADFLDLSRHRYALVAGRESSGGLISPPORALCILE--- 253  
QY 54 ELTEDYFSPVDNAELVLTILGQAMOGVSDIRRLV-SAFEPQVGLIQAAQOL----- 106  
DB 254 -----LTL-----EHCRRFCWSRHDKAISAVEKASLYLNTNL 287

```

OY 107 -----LCD-----EQAPORRLADLIHNVSINIAETRAEDPPWEGLESROSK 152
D 288 APSLQOLQVGLKYLQVEBEPQ---AAVAKLIRKASAVLSKSMKSPSP---LALALYESC 340
OY 153 SGYLRCSESRIRSEYLRVSSYSTVGAQOEFLRVIGSCQRLSMQVNGSFPDNGAK 212
D 341 QFLLSGLEGRKTRK---RDLALSLFAPFLGVCYCSLQQLQKRDGVY---GGS 386
OY 213 GSNLCTPEGWFCQGFPMDSCLSRHSINPYSNRESRIILFTWNLDHIIEKKRTIIPYL 272
D 387 SKQOQSFQWYF--QG-----LHLYTVVYDFAQGOQIVDLADLQGLVDSKSTYWM 437
OY 273 VEAKEDQGRVMEYRYGGLFTSENKLVIYCHKTKTKTKNDPSPRIYR 324
D 438 LEALEGLSGOELTDHMGMTASTYSN-----LAYSFSKRLVAEACALSEP 482

RESULT 18
US-09-867-550-1316
Sequence 1316, Application US/09867550
GENERAL INFORMATION:
APPLICANT: Leach, Martin D.
APPLICANT: Mehrahan, Foad,
APPLICANT: Conley, Pamela
APPLICANT: Law, Debbie
APPLICANT: Topper, James
TITLE OF INVENTION: No. US20020082206A1 Polynucleotides from Atherogenic Cells and
FILE REFERENCE: 21402-013 (Cura-313)
CURRENT APPLICATION NUMBER: US/09/867, 550
CURRENT FILING DATE: 2001-09-20
PRIOR APPLICATION NUMBER: USN 60/208,427
PRIOR FILING DATE: 2000-05-30
NUMBER OF SEQ ID NOS: 2125
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1316
LENGTH: 277
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)
OTHER INFORMATION: wherein Xaa may be any one of Ieu or Met or Val
US-09-867-550-1316

Query Match 4.6%; Score 82.5; DB 10; Length 277;
Best Local Similarity 23.8%; Pred. No. 11;
Matches 45; Conservative 20; Mismatches 51; Indels 73; Gaps 8;

OY 165 RSYLREYSSYSTVGAA-----QEFLLVIGSMQORLSMQVNGSFPDNGAKGSR 217
D 20 RRCIRLAAAMAALAAVAASRLERQEDIRFWLSEVERLADQLNSPY-----68
OY 218 CTBEGWSCQ--GPFMDSCLSRHSINPYSNRESRIILFTWNLDHIIT-----EK 264
D 69 -----SCQAEGR-----CLTR-----EVAQLRAEACDLRRLCSLRLCLAEBSR 108
OY 265 KRTIIPLVEAIKRODR-----EVDMEYVYGLFTSENKLVIHY 305
D 109 QATLESLELPAADGAGQPPPSQSDKDMKKKKKESADSEYKHOPITFERKLKLEIL 168
OY 306 CHKTKTKL 314
D 169 ---KKDHQL 174

RESULT 19
US-10-097-340-23
Sequence 23, Application US/10097340
GENERAL INFORMATION:
APPLICANT: John MONAHAN

```

```

APPLICANT: Manjula GANNAVARAPU
APPLICANT: Sebastian HOESCH
APPLICANT: Shubhangl KAMATKAR
APPLICANT: Steve G. KOVATS
APPLICANT: Rachel E. MEYERS
APPLICANT: Michael MORRISSEY
APPLICANT: Peter OLANDT
APPLICANT: Ami SEN
APPLICANT: Peter VEIRY
APPLICANT: Gordon B. MILLS
APPLICANT: Robert C. BAST, JR.
APPLICANT: Karen LU
APPLICANT: Rosemarie SCHMADT
APPLICANT: Xunel ZHAO
APPLICANT: Karen GLATT
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins for The Identification,
FILE REFERENCE: MRI-030
CURRENT APPLICATION NUMBER: US/10/097, 340
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: 60/276, 025
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/325, 149
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/276, 026
PRIOR FILING DATE: 2001-03-14
PRIOR APPLICATION NUMBER: 60/324, 967
PRIOR FILING DATE: 2001/09/26
PRIOR APPLICATION NUMBER: 60/311, 732
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: 60/325, 102
PRIOR FILING DATE: 2001-09-26
PRIOR APPLICATION NUMBER: 60/323, 580
PRIOR FILING DATE: 2001-09-19
NUMBER OF SEQ ID NOS: 363
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 633
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-340-23

Query Match 4.6%; Score 82.5; DB 9; Length 633;
Best Local Similarity 20.1%; Pred. No. 36;
Matches 73; Conservative 48; Mismatches 116; Indels 127; Gaps 18;

OY 37 FQLEGRSRLCLYEDGTETEDYFSPVDNAELVIL-----TLGQAMQCY 82
D 65 FSLNEGVRQLKTELGSEFFE-VLQNDLLTKGMVILRDKIRFEGQKLLDSLAETWDF 123
OY 83 SDIRRFSAFHEPQVGLQAAQQLCDEQAPORRLADLIH-----NVS-ONIAET 134
D 124 SDVLPMLQALIRPYQGRKPSVRLA-----LLHFNATITLSKLEDALARA 169
OY 135 RAEDPP-----WEGLESRFQSGYLRSCSRIRSYREVSYPSTVGAAQOE-- 184
D 170 HARVPALVQMLVLQGVHESRGVTEYLRL--LETLVQ--KVSPYLGTYGLHSSGPF 224
OY 185 -----EFLR-----VLGSMQORLSMQVNGS-----YFDGAKG-- 213
D 225 THSCITLQORKAAAALAVLGAVERKRPVPAAGDPALSTSHPYDVARHGIQLQVAGDR 284
OY 214 -GSRCTPEGWFC-----QGFPMDSCLSRHSINPYSNRESRIILFTWNLDHIIEKK 265
D 285 FGRYVIT-----FSCCRMPSPSHLDHQRILEYKTKTLQOYVENDTIYFHKGLN---SRN 337
OY 266 RTIIPLVEAIKEDGREVD-----WEYFYVGLFTSENKL 301
D 338 KPSLGLWLSAKYKEDRKQGDLTMPRLVYNSKLRSSHLSLPKYWDYR-----KKNLKA 392
OY 302 VHIY 305
D 393 LYVY 396

```

RESULT 20  
US-10-097-340-19  
Sequence 19, Application US/10097340  
Publication No. US20030087250A1  
GENERAL INFORMATION:  
APPLICANT: John MONAHAN  
APPLICANT: Manjula GANNAVARAPU  
APPLICANT: Sebastian HOERSCH  
APPLICANT: Shubhangi KAMATKAR  
APPLICANT: Steve G. KOVATS  
APPLICANT: Rachel E. MEYERS  
APPLICANT: Michael MORRISSEY  
APPLICANT: Peter OLANDT  
APPLICANT: Ami SEN  
APPLICANT: Peter VEIBY  
APPLICANT: Gordon B. MILLS  
APPLICANT: Robert C. BAST, JR.  
APPLICANT: Karen LU  
APPLICANT: Rosemarie SCHMANDT  
APPLICANT: Xumei ZHAO  
APPLICANT: Karen GIATT  
TITLE OF INVENTION: Nucleic Acid Molecules and Proteins For The Identification,  
TITLE OF INVENTION: Assessment, Prevention, and Therapy of Ovarian Cancer  
FILE REFERENCE: MRI-030  
CURRENT APPLICATION NUMBER: US/10/097,340  
PRIOR FILING DATE: 2002-03-14  
PRIOR APPLICATION NUMBER: 60/276,025  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/325,149  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/325,026  
PRIOR FILING DATE: 2001-03-14  
PRIOR APPLICATION NUMBER: 60/324,967  
PRIOR FILING DATE: 2001/09/26  
PRIOR APPLICATION NUMBER: 60/311,732  
PRIOR FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: 60/325,102  
PRIOR FILING DATE: 2001-09-26  
PRIOR APPLICATION NUMBER: 60/323,580  
PRIOR FILING DATE: 2001-09-19  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 19  
LENGTH: 718  
TYPE: PRT  
ORGANISM: Homo sapiens  
0-097-340-19

Query Match 4.68; Score 82.5; DB 9; Length 718;  
Best Local Similarity 20.18; Pred. No. 42; Indels 127; Gaps 18;  
Matches 73; Conservative 48; Mismatches 116;

QY 37 FOLPERGSRICLYEDGTETLTEDYFSPVDNAELVL-----TLQAMQGYV 82  
DB 65 FSLNEGVRQLKTELKLSFTE-YLQNOILLFGWVILRDKIRFEGOKLDSLAETWDFEF 123  
QY 83 SDIRRFLSAFHEPQVGLIDMAOQLCDEQAPQRORLADLLH-----NVS-QNTAAET 134  
DB 124 SDVLPMLQALFYVQCKEPEYVROLA-----LLHFRNAITLSVKLIEDALARA 169  
QY 135 RAEDDP-----WFGELSRFQSKGYLRYSCESTRISYLRVSSYSTVGAEOE-- 184  
DB 170 HARVPALVOMLIVLQGVHESRGVTEYLR--LETVYO--KVVSPLYGLYGHSEGEPE 224  
QY 185 -----EFLR-----VLGSMCORLRSMOYNGS-----YFDRGANG----- 213  
DB 225 THSCILQLQDKAAAVAGAVRKRPSVVPMAQOPALSTSHPPYDVARHGILQVAGDDR 284  
QY 214 -GSRICTPREGMFC-----QGFDMDSCL--RHSINPYSNRSRILFSTWNLDTIEKK 265  
DB 285 FGRVYV----FSCCMPPSHLDHQRLLELYKTYLIDQYVENDTIVYFHYGLN--SRN 337

QY 266 RTIPIYVAIKEDQGRVD-----WEYFGLTFSENKTL 301  
DB 338 KPSLGLQSAIYKEPFRKRGDILTMPRLVYSNKSIRKRSLSLTKYDYK-----KNLKA 392  
QY 302 VHTV 305  
DB 393 LYV 396

Search completed: May 26, 2003, 15:26:37  
Job time : 122 secs

**THIS PAGE BLANK (USPTO)**

Tue May 27 09:31:54 2003

us-09-748-451-2.ral

Page 1

Gencore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2003, 15:19:57 ; Search time 21 Seconds

(without alignments) 473.569 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1789

Sequence: 1 MLOKPSVKLRALRSPKRFQ.....SRVYKPTLRKQPVKRRQ 338

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database:

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/6C\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	1789	100.0	338	US-09-061-702-2	Sequence 2, Appl1
2	89	3.5	466	US-08-610-401-4	Sequence 4, Appl1
3	87	3.4	466	US-08-610-401-3	Sequence 3, Appl1
4	81	3.1	219	US-08-069-023-18	Sequence 18, Appl1
5	89	5.0	331	US-08-061-702-4	Sequence 4, Appl1
6	88.5	4.9	457	US-08-847-900-3	Sequence 3, Appl1
7	88	4.9	3080	US-09-085-199B-9	Sequence 9, Appl1
8	87	4.9	756	US-09-069-023-32	Sequence 32, Appl1
9	86.5	4.8	1463	US-08-157-005-3	Sequence 3, Appl1
10	86.5	4.8	1463	US-08-747-863-3	Sequence 3, Appl1
11	86.5	4.8	1463	US-09-565-864-3	Sequence 3, Appl1
12	86.5	4.8	219	US-09-069-023-12	Sequence 12, Appl1
13	86	4.8	2296	US-08-286-819A-27	Sequence 27, Appl1
14	85.5	4.8	2296	US-08-980-357-27	Sequence 27, Appl1
15	84.5	4.7	373	US-08-945-056-2	Sequence 16, Appl1
16	84	4.7	217	US-09-069-023-16	Sequence 16, Appl1
17	83.5	4.7	239	US-09-069-023-31	Sequence 31, Appl1
18	83.5	4.6	2618	US-09-413-814-28	Sequence 28, Appl1
19	82.5	4.6	299	US-09-069-023-20	Sequence 20, Appl1
20	82.5	4.6	1055	US-08-659-251-5	Sequence 5, Appl1
21	82.5	4.6	1055	US-08-256-490-5	Sequence 5, Appl1
22	82.5	4.6	1055	US-08-394-912A-6	Sequence 6, Appl1
23	82.5	4.6	1055	US-08-394-912A-6	Sequence 6, Appl1
24	81.5	4.6	628	US-08-394-912A-6	Sequence 6, Appl1
25	81.5	4.6	628	US-08-394-912A-6	Sequence 6, Appl1
26	81.5	4.6	630	US-08-394-912A-6	Sequence 6, Appl1
27	81.5	4.6	630	US-08-394-912A-6	Sequence 6, Appl1

*Applicants only*

28	81.5	4.6	630	US-08-394-912A-5	Sequence 5, Appl1
29	81.5	4.6	630	US-09-333-636-5	Sequence 5, Appl1
30	81.5	4.6	870	US-09-172-339-8	Sequence 8, Appl1
31	81	4.5	2818	US-08-510-284-1	Sequence 1, Appl1
32	81	4.5	2818	US-08-411-389-2	Sequence 2, Appl1
33	81	4.5	2818	US-08-449-933-2	Sequence 2, Appl1
34	81	4.5	2818	US-07-966-049A-2	Sequence 2, Appl1
35	81	4.5	2818	US-09-542-331-2	Sequence 2, Appl1
36	81	4.5	1001	US-09-060-410-2	Sequence 2, Appl1
37	80.5	4.4	626	US-07-938-782A-2	Sequence 2, Appl1
38	79.5	4.4	626	US-08-630-524-2	Sequence 2, Appl1
39	79.5	4.4	626	PCT-US93-08131-2	Sequence 2, Appl1
40	79.5	4.4	323	US-09-619-353-6	Sequence 2, Appl1
41	79	4.4	384	US-08-707-793A-5	Sequence 5, Appl1
42	78.5	4.4	384	US-08-707-793A-5	Sequence 5, Appl1
43	78.5	4.4	402	US-08-553-703A-1	Sequence 1, Appl1
44	78.5	4.4	402	US-09-006-021-1	Sequence 1, Appl1
45	78.5	4.4	402	US-09-006-021-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1  
US-09-061-702-2  
Sequence 2, Application US/09061702  
Patent No. 6165737  
GENERAL INFORMATION:  
APPLICANT: Wang, Xiaodong  
INVENTOR: Liu, Xuesong  
TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN  
TITLE OF INVENTION: APOPTOSIS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/061,702  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: McMillian, Nabeela R.  
REGISTRATION NUMBER: P-43,363  
REFERENCE/DOCKET NUMBER: UTSD:546  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512)418-3000  
TELEFAX: (512)474-7577  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 338 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-09-061-702-2  
Query Match  
Best Local Similarity 100.0%: Score 1789; DB 4; Length 338;  
Matches 338; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
DB  
1 MLOKPSVKLRALRSPKRFQVAGRCQEVLRKGLRFPQRPERSRLCYEDGTETEDYF 60  
1 MLOKPSVKLRALRSPKRFQVAGRCQEVLRKGLRFPQRPERSRLCYEDGTETEDYF 60  
QY  
61 PSVPDAEALVLLTIGQANQGYVSDIRRLSAPHFPOVGLIOAQQLLCDEQAPORRLA 120

```

Db      61 PSVPDNEVLTLTGQAMQGVSDIRRFSLAFHEPOVGLIOAQQLLCDEQAPQORLLA 120
QY      121 DLHNVSONIAETRAEDPPFEEGLERFSQSGYLRSCESTRSYLREVSYPSTYGA 180
Db      121 DLHNVSONIAETRAEDPPFEEGLERFSQSGYLRSCESTRSYLREVSYPSTYGA 180
QY      181 EAOEELRVLSGMCQRLRSQYNGSYFDRGAKGSGRLCTPEGWFSQCGPEMDSCLSRHS 240
Db      181 EAOEELRVLSGMCQRLRSQYNGSYFDRGAKGSGRLCTPEGWFSQCGPEMDSCLSRHS 240
QY      241 INPVSNESTRLESTWMLDHIIEKKRTIIPLYEAIKEQDGRVDMETRYGILLFTSENK 300
Db      241 INPVSNESTRLESTWMLDHIIEKKRTIIPLYEAIKEQDGRVDMETRYGILLFTSENK 300
QY      301 LVHIVCHKTTTHKLNCDPSRIYKQPTRLKRPYRKHQ 338
Db      301 LVHIVCHKTTTHKLNCDPSRIYKQPTRLKRPYRKHQ 338

```

```

2
-610-401-4
Sequence 4, Application US/09610401
Patent No. 6417336

```

```

GENERAL INFORMATION:
APPLICANT: MORISHIMA, NO. 6417336uhitro,
APPLICANT: NAKANISHI, Keiko,
APPLICANT: SHIBATA, Takehiko
TITLE OF INVENTION: Antibody against cleavage product of vimentin
FILE REFERENCE: 522.1004
CURRENT APPLICATION NUMBER: US/09/610.401
CURRENT FILING DATE: 2000-07-05
PRIOR APPLICATION NUMBER: JP 11-193235
PRIOR FILING DATE: 1999-07-07
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4
LENGTH: 466
TYPE: PRT
ORGANISM: Mus sp.
US-09-610-401-4

```

```

Query Match      5.5%: Score 99; DB 4; Length 466;
Best Local Similarity 20.1%; Pred. No. 0.054;
Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

```

```

QY      27 QEVLRKGLRFOQLPERGSRCLYEDGETELTEDYPPVPDNEVLTLTGQAMQGVSDIR 86
Db      180 EDIMR---LREKLOEE---MLQREAEASTLQSFROVDNASTLARLDLERVESTLOEIEI- 231
QY      87 RFLSAFHEPOVGLIOA-----QQLCDEQAPQORLLADLHNVSONIAETRAEDPPW 141
Db      232 AFLKRLHEEIOELQAOIOEOHVQIDVDVSKPDLTALRDVROOY-ESVAKNKIQEAEEW 290
QY      142 F-----GLSRFSQSGYLRY-----SCESRIR-----S 166
Db      291 YKSFADLSEANNNNDALROAKQESNEYRQOVSLTCEVDALKTNESLERQKREEMEN 350
QY      167 YLREVSYPSTYGAEOEELRVLSGMCQRLRSQYNGSYFDRGAKGSGRLCTPEGWFSQ 226
Db      351 FAVEANAYODTIG-RLODEIOMKEEMARHLREYO-----DILLNV 389
QY      227 QGPEMDSCLSRHSINPVSNESTRIL-----LSTWMLDHI-----IEKKRTIIP 272
Db      390 KMALDIEAIYRKIL---EGESRISLPLPFSSILNRETNLDSLPIVDTHSKRTL---L 443
QY      273 VEAIKEDGREV 285.
Db      444 IKVETRDGOVIN 456

```

```

RESULT 3
US-09-610-401-3

```

```

Sequence 3, Application US/09610401
Patent No. 6417336
GENERAL INFORMATION:
APPLICANT: MORISHIMA, NO. 6417336uhitro,
APPLICANT: NAKANISHI, Keiko,
APPLICANT: SHIBATA, Takehiko
TITLE OF INVENTION: Antibody against cleavage product of vimentin
FILE REFERENCE: 522.1004
CURRENT APPLICATION NUMBER: US/09/610.401
CURRENT FILING DATE: 2000-07-05
PRIOR APPLICATION NUMBER: JP 11-193235
PRIOR FILING DATE: 1999-07-07
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 466
TYPE: PRT
ORGANISM: Homo sapiens
US-09-610-401-3

```

```

Query Match      5.4%: Score 97; DB 4; Length 466;
Best Local Similarity 20.8%; Pred. No. 0.088;
Matches 65; Conservative 46; Mismatches 112; Indels 90; Gaps 15;

```

```

QY      27 QEVLRKGLRFOQLPERGSRCLYEDGETELTEDYPPVPDNEVLTLTGQAMQGVSDIR 86
Db      180 EDIMR---LREKLOEE---MLQREAEASTLQSFROVDNASTLARLDLERVESTLOEIEI- 231
QY      87 RFLSAFHEPOVGLIOA-----QQLCDEQAPQORLLADLHNVSONIAETRAEDPPW 141
Db      232 AFLKRLHEEIOELQAOIOEOHVQIDVDVSKPDLTALRDVROOY-ESVAKNKIQEAEEW 290
QY      142 F-----GLSRFSQSGYLRY-----SCESRIR-----S 166
Db      291 YKSFADLSEANNNNDALROAKQESTERYRQOVSLTCEVDALKTNESLERQKREEMEN 350
QY      167 YLREVSYPSTYGAEOEELRVLSGMCQRLRSQYNGSYFDRGAKGSGRLCTPEGWFSQ 226
Db      351 FAVEANAYODTIG-RLODEIOMKEEMARHLREYO-----DILLNV 389
QY      227 QGPEMDSCLSRHSINPVSNESTRIL-----LSTWMLDHI-----IEKKRTIIP 272
Db      390 KMALDIEAIYRKIL---EGESRISLPLPFSSILNRETNLDSLPIVDTHSKRTF---L 443
QY      273 VEAIKEDGREV 285
Db      444 IKVETRDGOVIN 456

```

```

RESULT 4
US-09-069-023-18
Sequence 18, Application US/09069023A
Patent No. 6348573
GENERAL INFORMATION:
APPLICANT: Nunez, Gabriel
APPLICANT: Inohara, Naohito
APPLICANT: Koseki, Takeyoshi
TITLE OF INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
FILE REFERENCE: UM-03333
CURRENT APPLICATION NUMBER: US/09/069.023A
CURRENT FILING DATE: 1998-04-27
NUMBER OF SEQ ID NOS: 38
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 18
LENGTH: 219
TYPE: PRT
ORGANISM: Mus musculus
US-09-069-023-18

```

```

Query Match      5.1%: Score 91; DB 4; Length 219;
Best Local Similarity 25.8%; Pred. No. 0.12;
Matches 46; Conservative 21; Mismatches 71; Indels 40; Gaps 7;

```



Db 110 QVESHKOHAKRKAKEKADKELPLIPVEEYKADAKLVLOREKONITSLVINEOE 169  
QY 178 -VGAEOEELRLVLSGMOQLR-----SMOYNGSYFDGANGSGRL- 217  
Db 170 EIGAYDDELQSRVSNLEERLRACMOQLACGKLNGISDPYVTKSGSF-----GSMNT 223  
QY 218 --CTPEG----WFSCQCFPMDSCLSRHSINPYSN 246  
Db 224 DPLAPEGDNRYWY-----MDGYHNNRFRREYKS 251

## RESULT 7

5223423-4  
; Patent No. 5223423  
; APPLICANT: FRANCHINI, GENOVEPPA, WONG-STAL, FLOSSIE;  
; GALLO, ROBERT  
; TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT  
; HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/15Y  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/331.212  
; FILING DATE: 03-31-1989  
; ID NO: 4  
; LENGTH: 3080

5223423-4

Query Match 4.9%; Score 88.5; DB 6; Length 3080;  
Best Local Similarity 27.8%; Pred. No. 13;  
Matches 40; Conservative 20; Mismatches 55; Indels 29; Gaps 7;

QY 16 PRKGVAGRSQEVLRKGLRGLRQLPERSRLCTLEDGELTDFPSPDNAE-----LYL 71  
Db 861 PHPGLAKKRRTYLDVGDVAFSIP-----LTEDRQYAFPLPSV-NNAEFGKRYIY 912  
QY 72 LTLQANOG-----YVSDIRFLSAFHE--PQVGLIAQAQQLCDEQAPQROQL----- 118  
Db 913 KVLPGKMGKSPAIQYIMRQVLEPFRKANPDVITIVQYMDLLI--ASDRIDLEHDKVYL 969  
QY 119 -LADLLHNVSQNIATRAEDPPW 141  
Db 970 QLKELLNGLGFSTDEKFOKDPY 993

## RESULT 8

US-09-085-199B-9  
; Sequence 9, Application US/09085199B  
; Patent No. 6235879  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Hacken, Abigail  
; APPLICANT: Hug, A.H.M. Mahabul  
; APPLICANT: Chopra, Vikramjit Singh  
; APPLICANT: Kalchman, Michael  
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Oppedahl & Larson  
; STREET: PO Box 5270  
; CITY: Frisco  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80443-5270  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.44 Kb storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS DOS 5.0  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085,199B  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:

NAME: Larson, Marina T.  
REGISTRATION NUMBER: 32038  
REFERENCE/DOCKET NUMBER: UBC-P-0130S2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (970) 668-2050  
TELEFAX: (970) 668-2052  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 756  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: no  
ORIGINAL SOURCE:  
ORGANISM: mouse  
FEATURE:  
OTHER INFORMATION: Huntington-interacting protein

US-09-085-199B-9

Query Match 4.9%; Score 88; DB 4; Length 756;  
Best Local Similarity 22.4%; Pred. No. 1.7;  
Matches 72; Conservative 47; Mismatches 104; Indels 98; Gaps 15;

QY 22 AGRSQEVLRKGLRGLRQLPERSRLCTLEDGELTDFPSPDNAEVLTLTQANOGY 81  
Db 169 AQRSLTELRKAQANEQ--RYSK--LKEKYSLEVNADLLRNAB----- 210  
QY 82 VSDIRFLSAFHEPQVGLIAQAQQL-----CDEQAPQROBLADLLHNVSQNIATRA 136  
Db 211 --VTKQVSARQAQVLEERKELADSPAVSQOARKQOEDVLENLKHELAT--SRQ 266  
QY 137 EDPWFEGLESRFQSKGYLYRSCESRIRSYLRVSSYPSTVGAEOEF----- 186  
Db 267 ELQVLSNLEFSASQEKMLTQIAE-----LEKQCSLATAVAQREELSLALDQEST 320  
QY 187 -LRYLG---SMCOILRSQVNGSYFDGA-----KGSRLCTPEGFSCGPPD----- 231  
Db 321 QIKLAGAESMCQGVQKDKRTLLAGIRKAAREIQEALSQLEET-LLSCGSTDLHLK 379  
QY 232 ---MDCLSR-----HSINPYSNRSRLRSTWNLDHIIRKRTII 269  
Db 380 VSSVSCLEQLERKNGSYLACPEDISLHSHIT-----LALHLTGDVIGSATSL 430  
QY 270 -----PTVEAIKEDDRE 283  
Db 431 RAPEPADSLTEACR-QYGRE 450

## RESULT 9

US-09-069-023-32  
; Sequence 32, Application US/09069023A  
; Patent No. 6348573  
; GENERAL INFORMATION:  
; APPLICANT: Nunez, Gabriel  
; APPLICANT: Inohara, Naohiro  
; APPLICANT: Koseki, Takeyoshi  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS  
; FILE REFERENCE: UM-03333  
; CURRENT APPLICATION NUMBER: US/09/069,023A  
; CURRENT FILING DATE: 1998-04-27  
; NUMBER OF SEQ ID NOS: 38  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Mus musculus

US-09-069-023-32

Query Match 4.9%; Score 87; DB 4; Length 331;  
Best Local Similarity 22.4%; Pred. No. 0.62;  
Matches 81; Conservative 50; Mismatches 131; Indels 100; Gaps 18;

```

QY 5 PRSVKRLALR-----SPRKEGVAGRSQCEVLR-KGC--LRFOLPERSRLCTIYEDGT 53
DB 10 PSEGETRLKEDLRRNRSROHGVAA-SCLEDLRKADIIAIKSLTPVTLVAEDGT 68
QY 54 EL-TEDYPSVDNAEVLITLGCAMOGYSDIRRLPSAFHBPQGLLOAQOOLL-----C 108
DB 69 IYDDDYFLCPSNTRKFVALASNEKW-----AYNSDGTAMISQESFDVDET 116
QY 109 DQAPORORLADLLHNVSONIARETRAE-----DPPWFEGLSPQSKGLYSCSES- 162
DB 117 DSGAGLKKNNARQOLKEDLSIILSHEDLOMLVDRPCSDLAQE-----LNGSCATV 168
QY 163 -RIRSVLREYSSYPTVGAENOEFLFVLSGMRKRSQYSGSFEDGAGSRLLCTPE 221
DB 169 ORLQHTLQOVL-----QREVR-----QSKQLLQVLALEKESLSKOE 210
QY 222 GNFSCGPPD-----MOSCLSRHSINFSNRESRIJFSTWINDIIEKKRTIPLVEAI 276
DB 211 ---ESKAAGEEVDADVDTGISRET-----SSDVALLS---HTL-----TAL 245
QY 277 KQDGRVDMVEFYGLFTSENLKVHIVCHKTKTKLNCDSRIYKPKOTLRKRQPVKR 336
DB 246 RERQAPE-----LSLSQDELVTYKEDPKALAVANMDIKTETVQEAECERLALRL 297
QY 337 RQ 338
DB 298 QQ 299

RESULT 10
US-08-157-005-3
; Sequence 3, Application US/08157005
; Patent No. 5620691
; GENERAL INFORMATION:
; APPLICANT: Wensvoort, Gert
; APPLICANT: Terpstra, Catharinus
; APPLICANT: Pol, Johannes M
; APPLICANT: Moorman, Robertus J
; APPLICANT: Meulenber, Johanna J
; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/157,005
; FILING DATE: 26-NOV-1993
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91201398.4
; FILING DATE: 06-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92200781.0
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00096
; FILING DATE: 05-JUN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Morai, Thomas F
; REGISTRATION/DOCKET NUMBER: 16,579
; REFERENCE/DOCKET NUMBER: 44819
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550

```

```

; TELEFAX: (212) 315-1931
; TELEX: 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1463 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-157-005-3

Query Match 4.8%; Score 86.5; DB 1; Length 1463;
Best Local Similarity 22.1%; Pred. No. 6.8;
Matches 63; Conservative 41; Mismatches 96; Indels 85; Gaps 15;

QY 100 IOAAQOOLLDEQAPORORLADLLHNVSONIARETRAE-----EDPPWFEGL 145
DB 558 IEAGROLV-----PNDRILAAALVHMKAAQASAYASAAILMDSACIDHDEPWYEDL 612
QY 146 ---ESRPSKSGY-----LRYSCSRIRSYLREVSYSPTVGAENOEFLFVLG-SWQ 195
DB 613 ICGIARCARQDDYSFPGPAFFWSMWEKLRSH-NEGKKFRGICIDAKADYASAGLDLC- 670
QY 196 RLRSQYNGSYFDR-----GAKGSRLLCTPEGWFSCGPF-----DMSCLSRHSIN 242
DB 671 ---LFSHFHQCXYTLSCGHHAGSKES-----QCSPVGAGRSPDLAVLKQIPYK 719
QY 243 PYSNRESRIJFSTWIND-----HIEKKRTIPLVEAIKQDGRVDMVEFYGLF 294
DB 720 PPRVYIMKGVNKTTLALDQGRYGRGLVAVKRG-----AGNEVDLSQXQ-VVPLLP 772
QY 295 TSENKLVHIVCH-----KRTTKHLN--CDPSRIYKP 324
DB 773 TCDIMNVKAVCNVLSKFIYGPGRSKITWLSQVDDVITYP 817

RESULT 11
US-08-747-863-3
; Sequence 3, Application US/08747863
; Patent No. 6197310
; GENERAL INFORMATION:
; APPLICANT: Wensvoort, Gert
; APPLICANT: Terpstra, Catharinus
; APPLICANT: Pol, Johannes M
; APPLICANT: Moorman, Robertus J
; APPLICANT: Meulenber, Johanna J
; TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Traak, Britt & Rosaa
; STREET: 525 South 300 East
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: WINDOWS95
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/747,863
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/157,005
; FILING DATE: 26-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91201398.4
; FILING DATE: 06-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 92200781.0
; FILING DATE: 18-MAR-1992

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

```

? APPLICATION NUMBER: 08/157,005
? FILING DATE: <unknown>
? APPLICATION NUMBER: EP 91201398.4
? FILING DATE: 06-JUN-1991
? APPLICATION NUMBER: EP 932200781.0
? FILING DATE: 18-MAR-1992
? APPLICATION NUMBER: PCT/NL92/00096
? FILING DATE: 05-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Moran, Thomas F
REGISTRATION NUMBER: 16,579
REFERENCE/DOCKET NUMBER: 44819
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEX: (212) 315-1931
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1463 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-565-864-3

Query Match          4.8%; Score 86.5; DB 4; Length 1463;
Best Local Similarity 22.1%; Pred. No. 6.8;
Matches    63; Conservative   41; Mismatches   96; Indels    85; Gaps    15;

OY IOAQQQLCDQAPQRRLDLHNVSQNIAETRA-----EDPPWFGL 145
Db IEAGQIL-----PVRRIILAALAHMAQNASEYYASAAMAILMDSACIDHDPWYBDL 612
OY ---BSRFQSKSGY-----LRSCESRIRSLREYSYPSTVGAEAOEFLRVLG-SMQC 195
Db ICGIARCARGDGGSPGAFPMSSMEKLRSH-NCKKKRRHGICDAKDADVASACLDC- 670
OY 196 RLRSQVNGSYFDR-----GAKGSRLCTPGSFGSQGF-----DMDCLSRHISIN 242
Db 671 -----LFHSHPHOHCXYTLCSGHNAGSKES----OCOSPVGAGRPLDALVKQIPYK 719
OY 243 PYNSRESRLISTWNLD-----HIIEKRRTIPTVEAIKEQDGREVDEMYEGLLF 294
Db 720 PPPIVIYMKGKTALTALDGRYQSRRLVAIVKRG-----AGNEVDLSDXDYQ-VPLLPL 772
OY 295 TSENKLTVHYCH-----KKTFRKLN--CDPSRIYKP 324
Db 773 TKCIIMNVKVCANVLISKFTYGPGSGSKTWTLSGYODDDVIYXP 817

RESULT 13
US-09-069-023-12
Sequence 12, Application US/09069023A
Patent No. 6348573
GENERAL INFORMATION:
APPLICANT: Nunez, Gabriel
APPLICANT: Inohara, Naohiro
APPLICANT: Koseki, Takeyoshi
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
TITLE OR INVENTION: SIGNALING PATHWAY INHIBITORS AND ACTIVATORS
FILE REFERENCE: UM-03333
CURRENT APPLICATION NUMBER: US/09/069,023A
CURRENT FILING DATE: 1998-04-27
NUMBER OF SEQ IDS NOS: 38
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 12
LENGTH: 219
TYPE: PROT
ORGANISM: Homo sapiens
US-09-069-023-12

Query Match      4.8%; Score 86; DB 4; Length 219;
Best Local Similarity 36.4%; Pred. No. 0.42;
```

Matches 24; Conservative 11; Mismatches 29; Indels 2; Gaps 2;  
QY 14 RSPKRGVAGRCQCVLAKGCLRPQLPERSRLCYEDTEL-TEDYPSYDPAALVYL 72  
DB 44 RSSRR-GVNASLDELISKILDAIVATGLVLYLEDDQTVDTLEFFQITLGDTHRMIL 102  
QY 73 TLGQAV 78  
DB 103 ERQOKW 108

RESULT 14  
US-08-286-819A-27  
; Sequence 27, Application US/08286819A  
; Patent No. 5871910

GENERAL INFORMATION:  
APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,819A  
FILING DATE: 05-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA: FR 9013579  
FILING DATE: 31-OCT-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5871910man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELETYPE: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2296 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-286-819A-27

Query Match 4.8%; Score 85.5; DB 2; Length 2296;  
Best Local Similarity 26.2%; Pred. No. 17;  
Matches 33; Conservative 17; Mismatches 41; Indels 35; Gaps 6;  
QY 156 LRYSCSS-RIRSEYLAEEVSSYSTVGAAQEEFLRVGSMCQRLRSMQV---NSGY----206  
DB 1314 LRFMCKSRIRIRITLISATKKY-----RSKREHFLSLIGAV--HLIPMSWLKHKQGNVAVPH 1366  
QY 207 --FDKAGKSGRLCTPECFSCQGFMDSCSLRHSINPYSNRESRIILFTWNLDTIEK 264  
DB 1367 WMTWKRRKSFSTLIPKQMLINPYNFKRECLSHRIRP-----11PS 1409  
QY 265 KRTIIP 270  
DB 1410 KRCVIP 1415

RESULT 15  
US-08-980-357-27  
; Sequence 27, Application US/08980357  
; Patent No. 6013508

GENERAL INFORMATION:  
APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/980,357  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,819  
FILING DATE: 05-AUG-1994  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
PRIOR APPLICATION DATA: FR 9013579  
FILING DATE: 31-OCT-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 6013508man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELETYPE: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 27:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2296 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein.  
US-08-980-357-27

Query Match 4.8%; Score 85.5; DB 3; Length 2296;

Best Local Similarity 26.2%; Pred. No. 17;

Matches 33; Conservative 17; Mismatches 41; Indels 35; Gaps 6;

156 LRYSES-RIRSYLREVSSYSTVGAEAEFLRYLGSRCORLSMOY-----NSGY----- 206

1314 LRFRCRSRIRITISATRYK-----RESKEHFLSLGAV--HLIPMSLKHKTGMWAVPH 1366

207 --FDGANGSRICLTPGCFSCQGFPMDSCLSRHSINPYSNRESRIILFTWNLDHIEK 264

1367 WMYKERSFSTLAPKKNLIINFLNFKECTLSHRIRP-----ILPS 1409

265 KRITIP 270

1410 KRCVIP 1415

RESULT 16

US-08-945-056-2

Sequence 2, Application US/08945056

Patent No. 6077994

GENERAL INFORMATION:

APPLICANT: Coupland, George M.

APPLICANT: Puttill, Joanna J.

TITLE OF INVENTION: Genetic control of flowering

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon & Vanderhye PC

STREET: 8th floor, 1100 No. 6077994th Glebe Road

CITY: Arlington

STATE: Virginia

COUNTRY: United States of America

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/945,056

FILING DATE: 20-OCT-1997

CLASSIFICATION: 800

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/G995/02561

FILING DATE: 01-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9422083.7

FILING DATE: 02-NOV-1994

ATTORNEY/AGENT INFORMATION:

NAME: Ms Mary J Wilson

REGISTRATION NUMBER: 32,955

REFERENCE/DOCKET NUMBER: 620-17

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 816-4000

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 373 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-945-056-2

Query Match 4.7%; Score 84.5; DB 3; Length 373;

Best Local Similarity 22.6%; Pred. No. 1.4;

Matches 47; Conservative 26; Mismatches 64; Indels 71; Gaps 10;

153 SGYLRSCSRIRIRYLREVSSYSTVGAEAEFLRYLGSRCORLSMOYGSYFDGRK 212

36 SAVLCMSCDQVHNSANRYAS-----RHKRVCS-CEAPAP----- 71

QY 213 GGSRLCTPEGWFCQGFPMDSCLSRHSINPYSNRESRI-----LFSTWNLDHIE 263

DB 72 --AFLEADSDSLC-----TACSEVHSANPLARHQRVILIPSGSFSMTTHQS 123

QY 264 KRITIPPLVEAIKQDREVD-----WEY-----FYGLFTSENIKLV--- 302

DB 124 EKTWDPKRLVVDDEEGEDDKAEVSWLFPNSDKNNNNQNGLLFSDYLNLVDYN 183

QY 303 ---HIVCHKTKTHLNDPSRIYKPR 326

DB 184 SMDYKFTGEYSQHONCS-----VPR 206

RESULT 17

US-09-069-023-16

Sequence 16, Application US/09069023A

Patent No. 6348573

GENERAL INFORMATION:

APPLICANT: Nunez, Gabriel

APPLICANT: Inohara, Naohiro

APPLICANT: Koseki, Takeyoshi

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS

FILE REFERENCE: UM-03333

CURRENT APPLICATION NUMBER: US/09/069,023A

CURRENT FILING DATE: 1998-04-27

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 16

TYPE: PRT

LENGTH: 217

ORGANISM: Mus musculus

US-09-069-023-16

Query Match 4.7%; Score 84; DB 4; Length 217;

Best Local Similarity 36.4%; Pred. No. 0.68;

Matches 24; Conservative 11; Mismatches 29; Indels 2; Gaps 2;

QY 14 RSPRFVAGRSCEVLRKGLRQLPERGSRICLYEGTEL-TEDYPSVDAEALVL 72

DB 44 RSSR-GYMASSLDLISKTIDVLYITGLVTLVEDGIVADTEEPOTLNDTHFWIL 102

QY 73 TLQAW 78

DB 103 EKGQW 108

RESULT 18

US-09-069-023-31

Sequence 31, Application US/09069023A

Patent No. 6348573

GENERAL INFORMATION:

APPLICANT: Nunez, Gabriel

APPLICANT: Inohara, Naohiro

APPLICANT: Koseki, Takeyoshi

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS

FILE REFERENCE: UM-03333

CURRENT APPLICATION NUMBER: US/09/069,023A

CURRENT FILING DATE: 1998-04-27

NUMBER OF SEQ ID NOS: 38

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 31

TYPE: PRT

LENGTH: 239

ORGANISM: Mus musculus

US-09-069-023-31

Query Match 4.7%; Score 83.5; DB 4; Length 239;

Best Local Similarity 28.8%; Pred. No. 0.89;

Matches 23; Conservative 19; Mismatches 35; Indels 3; Gaps 2;

QY 3 ORKSVKLRALRSRPF--GVAGRSQEVLRKGCILRFQLPERSRLCLYEDGTETL-TEDY 59  
Db 38 ETPRRAPCRVSTABRKVKGMASHLEIDLNVODILKPKPSVLIEDGTIVETEY 97  
QY 60 PPSVDPNAELVLTILGQAW 79  
Db 98 FOALAKDTMFVLLKQKWK 117

## RESULT 19

US-09-413-814-28  
Sequence 28, Application US/09413814

Patent No. 6225064

GENERAL INFORMATION:

APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH

APPLICANT: Bristol-Myers Squibb, CO.

APPLICANT: Beyer, Stefan

APPLICANT: Bloeker, Helmut

APPLICANT: Brandt, Petra

APPLICANT: Cino, Paul M

APPLICANT: Dougherty, Brian A

APPLICANT: Goldberg, Steven L

APPLICANT: Hotte, Gerhard

APPLICANT: Mueller, Joachim

APPLICANT: Reichenbach, Hans

TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polypeptide or

FILE REFERENCE: PCT/US 99/23535

CURRENT APPLICATION NUMBER: US/09/413,814

EARLIER FILING DATE: 1999-10-07

EARLIER APPLICATION NUMBER: DE 198 46 493.2

NUMBER OF SEQ ID NOS: 107

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 28

LENGTH: 2618

TYPE: PRT

ORGANISM: Sorangium cellulosum

US-09-413-814-28

Query Match 4.7% Score 83.5; DB 4; Length 2618;

Best Local Similarity 23.0%; Pred. No. 35;

Matches 54; Conservative 25; Mismatches 89; Indels 67; Gaps 7;

QY 40 PERGSRCLYEDGTETEDYPPSPVDNAE-----LYLTLTGQAWOGYVDIRFLSA 91

Db 2062 PERDPLARLHETELGPHLLALRQLAEVGAPEVRLAVTTGLAIGSESLR----- 2116

QY 92 FHEPQVGLQAAQOLICDEQAFQRORLIADILHNVSONIAETRAEDPPWFEGLE----- 146

Db 2117 ---PEVGLLGPVRYLPEEPNLRLLT-----DIDSADPTWRSQCEPLIRE 2160

QY 147 -----SRQSKSGYLRYSCESRIKSYLREVSYPSTVGAEOEFLRYLG 191

Db 2161 MGAAPGPEETALNGTSRWELGYPEGGTSTISSRLREGVYLITG-----LGLIG 2213

QY 192 SMCQRLSMQYNGSYDNGAGSRLCTPEGMFCQGP-----DMDCLISR 238

Db 2214 LALARHLARKRYATLILAGRGAAP---ARELHQAPAEFVPAATAIQMECCGAR 2265

## RESULT 20

US-09-069-023-20  
Sequence 20, Application US/09069023A

Patent No. 6348573

GENERAL INFORMATION:

APPLICANT: Nunez, Gabriel

APPLICANT: Inohara, Naohiro

APPLICANT: Koseki, Takayoshi

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS

FILE REFERENCE: US-03333

CURRENT APPLICATION NUMBER: US/09/069,023A

CURRENT FILING DATE: 1998-04-27  
NUMBER OF SEQ ID NOS: 38  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 20  
LENGTH: 299  
TYPE: PRT  
ORGANISM: Drosophila melanogaster  
US-09-069-023-20

## Query Match

Best Local Similarity 4.6%; Score 82.5; DB 4; Length 299;

Matches 22; Conservative 16; Mismatches 35; Indels 1; Gaps 1;

QY 6 KSVKLRALRSRPF--GVAGRSQEVLRKGCILRFQLPERSRLCLYEDGTETL-TEDYPPSPV 64

Db 12 KPEKVDVYRNIRIKKAVCASLSEIRSKVAKEFKCDHPTIILDSGTEIDDEYFPTLD 71

QY 65 DNAELVLTILGQAW 78

Db 72 ENTELVAVPEGEHW 85

Search completed: May 26, 2003, 15:24:23  
Job time : 23 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2003, 15:14:07 ; Search time 37 Seconds  
(without alignments)  
878.201 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1789

Sequence: 1 MGRKSVKRLALRSPKRF.....SRVYQRLRKQPVRRKQ 338

Listing table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	108.5	6.1	1920	2 A53188	perletrlin - mouse
2	105.5	5.9	571	2 T02213	NBS-LRR type resist
3	104	5.8	1778	2 T070382	apollipoprotein B -
4	99.5	5.6	2629	2 I46569	apollipoprotein B -
5	99	5.5	466	2 A45074	vimentin - human
6	99	5.5	466	2 A43803	vimentin - mouse
7	98.5	5.5	684	2 AF0627	helicase IV (75 kd
8	98	5.5	448	2 I48128	vimentin - Chinese
9	98	5.5	464	1 VEHY	vimentin - golden
10	98	5.5	466	2 S22119	slp(w7) - mouse (f
11	98	5.5	594	2 I48714	acetolactate synth
12	96.5	5.4	571	2 AG3174	COI intron 10 prot
13	93.5	5.2	397	2 C38888	HIV-1 retropepsin
14	92.5	5.2	1054	1 GNLJG5	sex-limited protei
15	92	5.1	1735	2 A54784	sex-limited protei
16	92	5.1	1736	2 A29176	oxido/reductase ir
17	91.5	5.1	925	2 D70441	ct234 hypothetical
18	91.5	5.1	925	2 A72096	conserved hypotet
19	91.5	5.1	925	2 B81573	CT234 hypothetical
20	91.5	5.1	925	2 B86527	vimentin - chicken
21	91	5.1	460	2 A29329	hypothetical prote
22	91	5.1	776	2 C87555	hypothetical prote
23	90	5.0	1044	2 B84497	hypothetical prote
24	90	5.0	1211	2 T41702	probable calson-tr
25	89.5	5.0	763	2 A47563	glucose-6-phosphat
26	89.5	5.0	416	2 T13719	calo protein - fru
27	89	5.0	457	2 T13636	neuronal olfactome
28	89	5.0	609	2 T28896	hypothetical prote
29	88.5	4.9	458	2 A43554	desmin - African c

30	88.5	4.9	1032	2 S12153	pol polyprotein -
31	88.5	4.9	1306	2 T13592	hypothetical prote
32	88.5	4.9	1738	1 A24558	complement C4 prec
33	87.5	4.9	1035	1 GNLJG6	HIV-1 retropepsin
34	87.5	4.9	1242	2 S78061	DNA-directed RNA p
35	87	4.9	1066	2 T45283	growth polarity ma
36	87	4.9	1066	2 T41099	staurosporine targ
37	86.5	4.8	1121	2 T38127	phosphoprotein pho
38	86	4.8	457	2 C75632	probable hemolysin
39	86	4.8	438	2 A43549	vimentin 1 - Affic
40	86	4.8	1027	2 B43549	vimentin 4 - Affic
41	86	4.8	1027	2 S37711	kinase heavy chai
42	86	4.8	1571	2 T14155	zinc finger protei
43	86	4.8	2895	2 H83362	hypothetical prote
44	85.5	4.8	987	2 T51360	kinase-like heavy
45	85.5	4.8	1744	1 C4H0	complement C4a pre

#### ALIGNMENTS

##### RESULT 1

A53188  
perletrlin - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 05-Nov-1999  
C:Accession: A53188  
R:Doxsey, S.J.; Stein, P.; Evans, L.; Calarco, P.D.; Kirschner, M.  
Cell 76, 639-650, 1994  
A:Title: Perletrlin, a highly conserved centrosome protein involved in microtubule o  
A:Reference number: A53188; MID:94170365; PMID:8124707  
A:Accession: A53188  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1920 <BOX>  
A:Cross-references: GB:005823; NID:9458667; PID:AA17886.1; PID:9458668  
C:Keywords: coiled coil

##### Query Match

Best Local Similarity 23.9%; Pred. No. 1.7; Length 1920;  
Matches 56; Conservative 45; Mismatches 90; Indels 43; Gaps 11;

QY	2 LQKPKSVKRLALRSPKRFVAGRSQEVY-----RKGC-----LRPQLPGRSGRICTY--ED 51	
DB	142 LQKPKSVKRLALRSPKRFVAGRSQEVY-----LRGRRAQETALLQSRQCELELRHOHARREEMALRSQOE 197	
QY	52 GTLETFEDYFSPVDNALVLTITLQAMOG---YVSDIRRLSAFHEPQVGLIAAOQL 107	
DB	198 AAETLEKRLRSEMEKNAQ-TIETLKQDWSERELCTENLRQELSLKHQSEMEGLQSQFOKE 256	
QY	108 CDEQAPQORRL-----ADLIHNVSONIA-----ERRADPPMFEGLSRFQSKG 154	
DB	257 LSEQKVELEKIFQAKHEAVSLKNEAQHOAAIKKIQEDIQSEHCQYLDLPQKPREREK 316	
QY	155 YLRSCSRIRSYLRVSSYPSTVGAEQAEFLVLSMCQRLRS--MOYNGSY 206	
DB	317 AKLEETLTQASY-----EDLKAQGEQEE-ILLNLSQLESMKTRRELNSW 361	

##### RESULT 2

T02213  
NBS-LRR type resistance protein - rice (fragment)  
C:Species: Oryza sativa (rice)  
C:Date: 05-Mar-1999 #sequence\_revision 05-Mar-1999 #text\_change 17-Nov-2000  
C:Accession: T02213  
R:Heister, D.; Kuhl, J.; Laurie, D.A.; Yano, M.; Sasaki, T.; Devos, K.; Graner, A.;  
Proc. Natl. Acad. Sci. U.S.A. 95, 370-375, 1998  
A:Title: Rapid reorganization of resistance gene homologues in cereal genomes.  
A:Reference number: Z14623; MID:98081880; PMID:9419382  
A:Accession: T02213  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-571 <LEI>

A:Cross-references: EMBL:AF032688; NID:g2792219; PIDN:AA06985.1; PID:g2792220  
 C:Genetics:  
 A:Gene: r1  
 C:Superfamily: disease resistance protein RPS2; leucine-rich alpha-2-glycoprotein repeat

Query Match 5.9%; Score 105.5; DB 2; Length 571;

Best Local Similarity 20.7%; Pred. No. 0.62; Mismatches 97; Indels 129; Gaps 13;

Matches 69; Conservative 38; Mismatches 97; Indels 129; Gaps 13;

QY 43 GSRLCYEDGTELEDPFVS-----PDNAELVLTGLQAMOGVSDIR--FLSA 91

DB 240 GSLLC-----TKDFEDDKNVLRSIEWELPSDKNNILPALRLSYHILPAILKRCFAFCSV 294

QY 92 FHEPOV-----GLIQAAQQLCDE-----QAPORQLADLL 123

DB 295 FHKYVPEKETLVQIMALGFIOSEFRTIEELSSYDELGRSEFHGHGQYMHAM 354

QY 124 HNVSNNIAE--TRAEDPWFEGLESFRQSGYRSCSRIHSRYLAEVSPSTGAE 181

DB 355 HDLAOSVMDCLRLDDP-----NSSSTRSSRHLSFSCHNKSR----- 395

QY 182 AQEELRYLGSRCORLRSQYNGSYFDGANGSRLCTPEGWFCGPFDMSCLSRSI 241

DB 396 SEDEFL-----GPKKARTLLLLNGYKSRTPSPD----- 425

QY 242 NPYSNRESRIIFSTWNLDHIEKKRTIIPVLEAIKEDGREDVMEF----- 289

DB 426 -----FLMLRLYHVELNRRDITELPDSI-----GNLKMRLYLNLSGTGITVLP 471

QY 290 -YGLFTSENKLVHIVCHKKTHKLNCDPSRI 321

DB 472 SIGRLFNQITLKL-----KNCVLEICIPSI 497

### RESULT 3

UT0382

apolipoprotein B - p1g (fragments)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 31-Mar-1992 #sequence\_revision 31-Mar-1992 #text\_change 13-Aug-1999

C:Accession: UT0382; I46567; I46568

R:Maeda, N.; Ebert, D.L.; Doers, T.M.; Newman, M.; Hasler-Rapacz, J.; Attie, A.D.; Rapacz, Gene 70, 213-229, 1988

A:Title: Molecular genetics of the apolipoprotein B gene in pigs in relation to atherosclerosis

A:Reference number: UT0382; MUID:89100006; PMID:2905687

A:Accession: UT0382

A:Molecule type: DNA

A:Residues: 1-1778 <MAE>

R:Maeda, N.; Ebert, D.L.; Doers, T.M.; Newman, M.; Hasler-Rapacz, J.O.; Attie, A.D.; Rapacz, Gene 69, 213-229, 1988

A:Title: Molecular genetics of the apolipoprotein B gene in pigs in relation to atherosclerosis

A:Reference number: I46567

A:Accession: I46567

cus: translated from GB/EMBL/DBD

Residues: 1-8, 'S', 10-238 <MA2>

A:Molecule type: DNA

A:Cross-references: GB:M2666; NID:g164366; PIDN:AAA30996.1; PID:g164366

A:Accession: I46568

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 239-1778 <MA3>

A:Cross-references: GB:M2667; NID:g164367; PIDN:AAA30997.1; PID:g164370

C:Comment: Apolipoprotein B is the predominant protein component of the low-density lipoprotein

C:Genetics:

A:Gene: apob

A:Introns: 39/3; 88/3; 159/2; 238/3; 1133/1; 1171/2; 1232/3

A>Note: the list of introns may be incomplete

C:Superfamily: apolipoprotein B

C:Keywords: atherosclerosis; cholesterol metabolism; LDL; lipid binding; lipoprotein

Query Match 5.8%; Score 104; DB 2; Length 1778;

Best Local Similarity 19.5%; Pred. No. 3.8; Matches 66; Conservative 52; Mismatches 123; Indels 98; Gaps 14;

QY 36 RFOLEP-----GSRIC-----LY-----EDGTELEDPFVSPPNAEL----- 69

DB 1404 RFOLEGRANNTGDELCKMVMTEVEGLVSLQYKSHSGLIELLSYFODLMKSKLNALK 1463

QY 70 -----VLTLGAMOGVSDIRRFSAFHEPOVGLQAAQQLC----- 108

DB 1464 IKTFPDSKYQLTDVSEYGEQLKSLSDVQALSDLSINITTELSLQIFLECFQEI 1523

QY 109 -DEQAPORQLADLHNVSNNIAETRAEDPWFEGLESFRQSGYRSCSRI-- 164

DB 1524 EEEFLRLKEKRLADPINDIQNINNTFNIVAPLGRRLKENLDSPFGLNFIQNTLMEA 1583

QY 165 RSYLREVSST-----PSYVG-----AAQEEFLRYGSMCQRLRSQYNGSYFDR 209

DB 1584 SOELQOLHQYIKALKEFEYDPSMGWTVKYYELEKEKYNILIKNLVDYK--DFHSKYTVS 1641

QY 210 GAKGSRCTPEGWFCGPFDMSCLS-----RHSINPYSNRESRIIFSTWNL 260

DB 1642 ATDFASQLSQVOEFOVE--DIQYLSITLADADGKGKELAEUSRAQETI--KSAV-- 1695

QY 261 IIEKKRTIIPVLEAIKEDGREDVMEFYGLFTSEN 299

DB 1696 -----TMKEITISYHQR-----FTYKLDPDQL 1719

### RESULT 4

I46569

apolipoprotein B - p1g (fragment)

C:Species: Sus scrofa domestica (domestic pig)

C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 13-Aug-1999

C:Accession: I46569

R:Puttelli, C.; Maeda, N.; Ebert, D.L.; Kaiser, M.; Lund-Katz, S.; Sturley, S.L.; Kodo, J. Lipid Res. 34, 1323-1335, 1993

A:Title: Nucleotide sequence encoding the carboxyl-terminal half of apolipoprotein B

A:Reference number: I46569; MUID:94014802; PMID:8409766

A:Accession: I46569

A>Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-2629 <PDR>

A:Cross-references: GB:L11235; NID:g164371; PIDN:AAA74655.1; PID:g951375

C:Genetics:

A:Gene: APOB

A:Introns: 1984/1; 2022/2; 2083/3

C:Superfamily: apolipoprotein B

Query Match 5.6%; Score 99.5; DB 2; Length 2629;

Best Local Similarity 18.7%; Pred. No. 15; Matches 72; Conservative 61; Mismatches 140; Indels 113; Gaps 16;

QY 36 RFOLEP-----GSRIC-----LY-----EDGTELEDPFVSPPNAEL----- 69

DB 2255 RFOLEGRANNTGDELCKMVMTEVEGLVSLQYKSHSGLIELLSYFODLMKSKLNARK 2314

QY 70 -----VLTLGAMOGVSDIRRFSAFHEPOVGLQAAQQLC----- 108

DB 2315 IKTFPDSKYQLTDVSEYGEQLKSLSDVQALSDLSINITTELSLQIFLECFQEI 2374

QY 109 -DEQAPORQLADLHNVSNNIAETRAEDPWFEGLESFRQSGYRSCSRI-- 164

DB 2375 EEEFLRLKEKRLADPINDIQNINNTFNIVAPLGRRLKENLDSPFGLNFIQNTLMEA 2434

QY 165 RSYLREVSST-----PSYVG-----AAQEEFLRYGSMCQRLRSQYNGSYFDR 209

DB 2435 SOELQOLHQYIKALKEFEYDPSMGWTVKYYELEKEKYNILIKNLVDYK--DFHSKYTVS 2492

QY 210 GAKGSRCTPEGWFCGPFDMSCLS-----RHSINPYSNRESRIIFSTWNL 260

DB 2493 ATDFASQLSQVOEFOVE--DIQYLSITLADADGKGKELAEUSRAQETI--KSAV-- 2546

QY 261 IIEKKRTIIPVLEAIKEDGREDVMEFYGLFTSENKLVHIVCHK-- 308

DB 2547 -----TMKEITISYHQR-----FTYKLDPDQL 2594

```

OY 309 -----KTHKLNCDPSRIYKPPRL 328
DB 2595 ELRYMKLKELOSDTVMDMRPIKV 2620

RESULT 5
A25074
vimentin - human
C:Species: Homo sapiens (man)
C>Date: 30-Jun-1988 #sequence_revision 10-May-1996 #text_change 13-Aug-1999
C:Accession: S13115; A25074; J02034; S08487; I54056
R:Honore, B.; Madsen, P.; Basse, B.; Andersen, A.; Walbum, E.; Celis, J.E.; Leffers, H.
Nucleic Acids Res. 18, 6692, 1990
A>Title: Nucleotide sequence of cDNA covering the complete coding part of the human vime
A:Reference number: S13115; M01D:91067467; PMID:2251132
A:Accession: S13115
Molecule type: mRNA
A:Residues: 1-466 <HON>
A:Cross-references: EMBL:X56134; NID:937849; P1DN:CAA39600.1; PID:937850
R:Ferrari, S.; Beldini, K.; Macnatek, L.; Rittling, S.; Calabretta, B.; De Riel, J.K.;
Mol. Cell. Biol. 6, 3614-3620, 1986
A>Title: Coding sequence and growth regulation of the human vimentin gene.
A:Reference number: A25074; M01D:87089701; PMID:3467175
A:Accession: A25074
Molecule type: DNA
A:Residues: 1-41, 'D', '43-441', 'F', '443-466 <FER>
A:Cross-references: GB:M14144; NID:9340218; P1DN:AAA61279.1; PID:9340219
R:Perréau, J.; Lillienbaum, A.; Vasseux, M.; Paulin, D.
Gene 62, 7-16, 1988
A>Title: Nucleotide sequence of the human vimentin gene and regulation of its transcript
A:Reference number: J02034; M01D:88226018; PMID:3371665
A:Accession: J02034
Molecule type: DNA
A:Residues: 113-200, 'S', '202-264', 'S', '266-277', 'I', '279-466 <PER>
A>Note: the authors translated the codon GCG for residue 287 as Asp
R:Sommers, C.L.; Walker-Jones, D.; Heckford, S.E.; Worland, P.; Valverius, E.; Clark, R.
Cancer Res. 49, 4258-4263, 1989
A>Title: Vimentin rather than keratin expression in some hormone-independent breast can
A:Reference number: S08487; M01D:89303836; PMID:2472876
A:Accession: S08487
A>Status: preliminary
Molecule type: mRNA
A:Residues: 1-113, 'P', '114-135 <SOM>
R:Gupta, A.K.; Audin, J.E.; Wayer, M.M.
Gene 86, 303-304, 1990
A>Title: Isolation of a human vimentin cDNA with a long 3'-noncoding region from a human
A:Reference number: I54056; M01D:90215314; PMID:2323579
A:Accession: I54056
Molecule type: translated from GB/EMBL/DDBT
A:Residues: 167-466 <RES>
A:Cross-references: GB:M25246; NID:9340233; P1DN:AAA61282.1; PID:9340234
C:Genetics:
A:Gene: GDB:VIM
A:Cross-references: GDB:119630; OMIM:193060
A:Map position: 10p13-10p13
A:Introns: 188/2; 208/3; 294/3; 336/3; 410/2; 425/1; 453/3
C:Superfamily: cytoskeletal keratin
C:Keywords: coiled coil; intermediate filament; phosphoprotein

Query Match 5.5%; Score 99; DB 2; Length 466;
Best Local Similarity 20.8%; Pred. 1.7;
Matches 65; Conservative 47; Mismatches 111; Indels 90; Gaps 15;

OY 27 OEYLKRGCLRPOLPRGSRNCLYEGTLEMDYPPSPVDAEILYLTTCGACGCVASIR 86
DB 180 EDIMR-----LREKLGEE-----MIGREAEMLTOSRQOVNDAISARLDLREKRESIQEEL- 231
OY 87 RELSAFHEPOVGLTQAA-----QQLLDEQAPQORQLLADLHNVSQNTAATRAEDPPW 141
DB 232 AFLKRIHEIEIQLAQIQEOHQVLDVDSKFDLITAAALRDVQRQY-ESVAANKNLOEAEEM 290

RESULT 6
A43803
vimentin - mouse
C:Species: Mus musculus (house mouse)
C>Date: 12-Feb-1993 #sequence_revision 12-Feb-1993 #text_change 20-Jun-2000
C:Accession: A43803; S32832; J50241; S12774; S14526; A31951; B31951; S32627
R:Capetanaki, Y.; Kuisik, I.; Rothblum, K.; Starnes, S.
Oncogene 5, 645-655, 1990
A>Title: Mouse vimentin: structural relationship to fos, jun, CREB and tpr.
A:Reference number: A43803; M01D:90265604; PMID:2140597
A:Accession: A43803
Molecule type: mRNA
A:Residues: 1-466 <CAP>
A:Cross-references: EMBL:X51438; NID:955290; P1DN:CAA5803.1; PID:955291
R:Rauscher, A.
submitted to the EMBL Data Library, April 1993
A:Description: Upstream region of the mouse vimentin gene.
A:Reference number: S32627
A:Accession: S32832
A>Status: preliminary
Molecule type: DNA
A:Residues: 1-70 <RAU>
A:Cross-references: EMBL:Z22526; NID:91518339; P1DN:CAA80251.1; PID:9293950
R:Wood, L.; Thorburn, N.; Vogel, G.
Gene 76, 171-175, 1989
A>Title: Vimentin cDNA clones covering the complete intermediate-filament protein are
A:Reference number: J50241; M01D:89306653; PMID:2744479
A:Accession: J50241
Molecule type: mRNA
A:Residues: 1-155, 'EL', '158-337', 'E', '339-373', 'D', '375-466 <MOO>
A:Cross-references: GB:M24849; NID:9202367; P1DN:AAA40555.1; PID:9202368
R:Hennekkes, H.; Kuehn, S.; Traub, P.
Mol. Gen. Genet. 221, 33-36, 1990
A>Title: Coding sequence and flanking regions of the mouse vimentin gene.
A:Reference number: S12774; M01D:90220517; PMID:2325630
A:Accession: S12774
A>Status: translation not shown
Molecule type: mRNA
A:Residues: 1-155, 'EL', '158-337', 'E', '339-466 <HEN>
A:Cross-references: EMBL:M26251; NID:9202369; P1DN:AAA40556.1; PID:9202370
R:Podolin, P.L.; Prystowsky, M.B.
submitted to the EMBL Data Library, October 1990
A:Description: Nucleotide sequence of a murine vimentin cDNA.
A:Reference number: S14526
A:Accession: S14526
Molecule type: mRNA
A:Residues: 1-155, 'EL', '158-163', 'E', '165-337', 'E', '339-466 <POD>
A:Cross-references: EMBL:X56397; NID:955407; P1DN:CAA39807.1; PID:955408
R:ando, S.; Tanabe, K.; Conda, Y.; Sato, C.; Inagaki, M.
Biochemistry 28, 2974-2979, 1989
A>Title: Domain- and sequence-specific phosphorylation of vimentin induces disassemb
A:Reference number: A31951; M01D:89302884; PMID:2500966
A:Accession: A31951
Molecule type: protein
A:Residues: 5-12; 14-28; 37-69 <AND>
A:Accession: B31951

```

A:Molecule type: protein  
 A:Residues: 5-12;14-69 <AN2>  
 C:Superfamily: cytoskeletal keratin  
 C:Keywords: coiled coil; intermediate filament; phosphoprotein  
 F:104-138/Region: coil 1A  
 F:147-247/Region: coil 1B  
 F:264-282/Region: coil 1B  
 F:291-411/Region: coil 2B  
 F:7,25,39,51,66/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase and F  
 F:9,10,21,26,34,42/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #statu  
 F:47/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status experim

Query Match 5.5%; Score 99; DB 2; Length 466;  
 Best Local Similarity 20.1%; Pred. No. 1.7; Mismatches 113; Indels 90; Gaps 14;

Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

QY 27 QEVLRKGLRQLPRLPERSRLCLYEDGETLTDYPPSPDNLVLTGQAGVSDIR 86  
 DB 180 EDIMR-----LRKLOEE-----MLQREASTIQSFROVDNASTARLDLKEVESIQEET- 231

87 RFLSAFHEPOVGLIOAA-----QQLCDEQAPQORLADLLHNVSONIAETRAEDPPW 141  
 232 AFLKKLHDEEIQELQAOIQEOHQVIDVDSKPDLTALRDVROQY-ESVAKNLQEAEEW 290

QY 142 F-----EGLESRFQSKGYLR-----YSCESRIR-----S 166

DB 291 YKSKFADLSEANRNNDLRQAKQESNRYRQVOSTLCEVDALKTNYSLERQKMEEN 350

QY 167 YLREVSSYPSTVGAEAEQEFRLVLSMCQRLRSMQYNSYDPRGAKGGSRLCTPEGWFC 226

DB 351 FALSAANYQDTIG-RLQDEIQNMKEEMARHLREYQ-----DLLNV 389.

QY 227 QGPFDMDSCLSRHSINPYSNRESRI-----LFTWNLDHT-----IEKKRTIIPTL 272

DB 390 KMALDIEIATYRKLL-----EGESRISLPLPFTSSILNRETMLSLPLVDTHSKRTL--L 443

QY 273 VEAIKEDGPREVD 285

DB 444 IKTVETRDGOVIN 456

RESULT 7

AF0627  
 heliase IV (75 kD heliase) [imported] - Salmonella enterica subsp. enterica serovar Ty  
 C:Species: Salmonella enterica subsp. enterica serovar Typh  
 A:Note: This species has also been called Salmonella typhi

C>Date: 09-Nov-2001 #sequence\_revision 09-Nov-2001 #text\_change 27-Nov-2001

C:Accession: AF0627

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,  
 th, T.; Connetton, P.; Croplin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,  
 S.; Moulé, S.; O'Garra, P.

Ref: 413, 848-852, 2001

Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;  
 le: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Accession: AF0627

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-684 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD08201.1; PID:G16502250; GSPDB:GN00176

C:Genetics:

A:Gene: STY1097

C:Superfamily: heliase IV

Query Match 5.5%; Score 98.5; DB 2; Length 684;

Best Local Similarity 22.9%; Pred. No. 3.1;

Matches 89; Conservative 52; Mismatches 148; Indels 99; Gaps 22;

QY 26 QEVLRKGLRQLPERSRL---CLYEDG-TELTDYPPSPVD-----NAEL 69

DB 152 CREITMR-CLAWLQDSESGRQHQAYADAMLEAHADFQIIESSPLNPSQARAVNGES 210

QY 70 VLITLQAGMAGVSDI---RFLSAFHEPOVGLI-----QAAQLCDEQAPQOR---Q 116

DB 211 SLVLVLAGSGSKTSVLVARACWLLARQADAGQITLLAFGRKAAEEM--DERIRERLATE 268  
 QY 117 RLADLLHNVSONIAETRAEDPPWEGLES-----RFQSKGYLRYSCESE-----IRS 166  
 DB 269 EITARTFISLALYIT-IOGSKKAPVSKLSEDAFARHQLFHTWRQOCSEKKAQAKGRQ 327  
 QY 167 YLREVSSYPSTVGAEAEQEFRLVLSMCQRLRSMQYNSYDPRGAKGGSRLCTPEGWFC 224  
 DB 328 WLEEMQWVVEGNEFWDEDTQRLRLAPLRDWRVSLMRHG-----GAQAEIMAGAPE--- 379  
 QY 225 SCQGFDMDSCLSRHSINPYSNRESRI-----LFTWNLDHT-----IEKKRTIIP- 270  
 DB 380 ECRLEFGRIKLMAPLAKMSALKAEAVDFS-GLTHQAMVILDEKGFISPMKHLIYD 437  
 QY 271 -----TLVEAIKEDGPREV-----DMEYFGILFTSENKLVHIV----- 305  
 DB 438 EFQDISPQRAALLERLKRQNSQITLFFANGDDMOAIT--RFGAQLSLTFAHQTGEGEH 495  
 QY 306 CHKTKRLNCDPSRIYKPTRLKRP 333  
 DB 496 CHLDTTYRFN---SRIGDIANRFVQNP 520

RESULT 8

148128  
 vimentin - Chinese hamster (fragment)

C:Species: Crictetus griseus (Chinese hamster)

C>Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 13-Aug-1999

C:Accession: 148128

R:Blomendaal, H.; Quax, W.; Quax-Jeuken, Y.; Dodefont, H.; Ramaekers, F.; Dunia, I.;  
 Mol. Biol. Rep. 9, 113-118, 1983

A:Title: Organization and expression of the vimentin gene.

A:Reference number: 148128; MUID:83297272; PMID:6688458

A:Accession: 148128

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-448 <RES>

A:Cross-references: GB:M16718; NID:G191232; PIDN:AAA37029.1; PID:G387058

C:Superfamily: cytoskeletal keratin

Query Match 5.5%; Score 98; DB 2; Length 448;

Best Local Similarity 20.1%; Pred. No. 2; Mismatches 113; Indels 90; Gaps 14;

Matches 63; Conservative 47; Mismatches 113; Indels 90; Gaps 14;

QY 27 QEVLRKGLRQLPRLPERSRLCLYEDGETLTDYPPSPDNLVLTGQAGVSDIR 86

DB 162 EDIIR---LRKLOEE-----MLQREASTIQSFROVDNASTARLDLKEVESIQEET- 213

QY 87 RFLSAFHEPOVGLIOAA-----QQLCDEQAPQORLADLLHNVSONIAETRAEDPPW 141

DB 214 AFLKKLHDEEIQELQAOIQEOHQVIDVDSKPDLTALRDVROQY-ESVAKNLQEAEEW 272

QY 142 F-----EGLESRFQSKGYLR-----YSCESRIR-----S 166

DB 273 YKSKFADLSEANRNNDLRQAKQESNRYRQVOSTLCEVDALKTNYSLERQKMEEN 332

QY 167 YLREVSSYPSTVGAEAEQEFRLVLSMCQRLRSMQYNSYDPRGAKGGSRLCTPEGWFC 226

DB 333 FALSAANYQDTIG-RLQDEIQNMKEEMARHLREYQ-----DLLNV 371

QY 227 QGPFDMDSCLSRHSINPYSNRESRI-----LFTWNLDHT-----IEKKRTIIPTL 272

DB 372 KMALDIEIATYRKLL-----EGESRISLPLPFTSSILNRETMLSLPLVDTHSKRTL--L 425

QY 273 VEAIKEDGPREVD 285

DB 426 IKTVETRDGOVIN 438

RESULT 9

VEHY

vimentin - golden hamster



Db 327 KLGLSGMAIDITLLSFFHRLGDLKLTSLSDRYVHFETDGPVILL-YFDSVPTTREC 385  
 QY 70 VLLTLGAMOGYVSDIRRFSAFHEPOVGLIOAAQQLCD-----EQAPORQL 118  
 Db 386 V-----GF-----GASGEVYVGLVQPSAVLYDYSPDHKCSVYAPATKSOL 428  
 QY 119 LADLHNVSQNIATAETAEDEPPWEGLESFQSKSGY-LRYSC-----160  
 Db 429 LATL---CSGDVCCAGCKPPLRLSLERREDKDGFRMPACVYPRVEGFTVKVLRED 485  
 QY 161 -----ESRISYLEVSSYPTVGAEQAEERLVLGSMCQRLRSMQVNGSFEDGAK 212  
 Db 486 GRAAFRLPEKTIQVLFRIIDTMASIC-----QTRNLSRTSKRLR-LDEPKETLLMGMD 539  
 QY 213 G 213  
 Db 540 G 540

## RESULT 12

lactate synthase [imported] - Agrobacterium tumefaciens (strain C58, Dupont) plasmid  
 Accession: Agrobacterium tumefaciens  
 Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 01-Feb-2002

C:Accession: AG3174  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCelli  
 science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A>Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; PMID:11743193  
 A:Accession: AG3174  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-571 <KUR>  
 A:Cross-references: GB:AE008687; PIDN:AA45813.1; PID:g17743551; GSPDB:GN00188  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:  
 A:Gene: atrc  
 A:Genome: plasmid  
 C:Superfamily: acetylactate synthase large chain; thiamin pyrophosphate-binding domain

Query Match 5.48; Score 96.5; DB 2; Length 571;  
 Best Local Similarity 21.68; Pred. No. 3.6;  
 Matches 79; Conservative 49; Mismatches 122; Indels 115; Gaps 20;

QY 10 LRALRSRKRKGVAGRCQEVLRKGLARFQRPERSRLCYEDGELTEDYPPSPDAEL 69  
 Db 259 LALVNGKSL--GRHVEIVREADLVLLIGTRTN-----QNGTDNRQYSP--DAOI 306  
 QY 70 VLLTLGAMOGYVSDIRRFSAFHEPOVGLIOAAQQLCDQAPORQL-LADL-LHNVS 127  
 Db 307 IHIDTDQEGVGRNFARLVDARETLAQL-----RRIRLCDLHRTVS 351  
 QY 128 QNIAETRAEDPPEEGLESFQSKSGY-LRYSCS-----RISYLEVSSYPTVGAE 181  
 Db 352 REAVCNLRRES--WRKFNDR--RGY--YSSEASPLRPRIWAELQGVIDENTITYAD 403  
 QY 182 AOEERFLNVLSM--CQRLRSMQVNGSYFPGANG-----GSLRLCPE-----221  
 Db 404 ASYSSMNVLLQRLISSERTVLT-----PRGLAGLQWVPLAIGAVVAESPDAVVG 457  
 QY 222 -GNFSCGPPDMSCLSRHSINPYSNRESHLESTW-NLDHIEKKRTIIPTEVLAIK-- 277  
 Db 458 DGEFA-----HS-----MALELV--RMQIPITIVVANG 486  
 QY 278 -----EDDREVDNEYGLFTSENLKLVHIVCKRTTHKLNCDPSRIYKQTRLRKRP 333  
 Db 487 ILGFORDAETVYKFGKTYTACHFAE-----VDHVKTAHACGDAVAVTQPODLKAMQK 539  
 QY 334 VKRRQ 338

Db 540 VKRRK 544

## RESULT 13

COI intron 10 protein - Podospira anserina mitochondrion  
 C:Species: mitochondrion Podospira anserina  
 C:Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 03-Aug-2001  
 C:Accession: C38888

R:Cummings, D.J.; Michel, F.; McNally, K.L.

curr. Genet. 16, 381-406, 1989  
 A>Title: DNA sequence analysis of the 24.5 kilobase pair cytochrome oxidase subunit I  
 A:Reference number: A48327; MUID:90124722; PMID:2558809

A:Accession: C38888  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-397 <COM>  
 A:Cross-references: GB:X5026; GB:M30937; GB:M61734; MID:g14030; PIDN:CAA38788.1; PID  
 C:Genetics:  
 A:Genome: mitochondrion  
 A:Gene: SGC3  
 C:Superfamily: Podospira anserina mitochondrion COI intron 10 protein; COI intron 9  
 C:Keywords: mitochondrion  
 F:157-380/Domain: COI intron 9 protein homology <C19>

Query Match 5.28; Score 93.5; DB 2; Length 397;  
 Best Local Similarity 22.48; Pred. No. 4;  
 Matches 64; Conservative 39; Mismatches 112; Indels 71; Gaps 12;

QY 45 RLCLYEDGTELT---EDYPPSPDNLVLLTLGAMOGYVSDIRRFSAFHEPOVGLI 100  
 Db 49 KILVIMNPQITTKARSENKPKGITEFGLCW-----VWGISEAIRLPLPSLSKRIIDL 102  
 QY 101 QAAQOL--LCDEQAPQORQLADLHNVSQNIATAETRAEDPPEEGLESFQ-----150  
 Db 103 FALAQYIRLEKGYFLMKNLKYTLASDYSENSLANTRNNNSPFKEKEDFNEMLAGIT 162  
 QY 151 -----SKSGYLYRSCESRIRS---YLREVSYPSTVGAEQAEERL-----188  
 Db 163 DGDGCFQVSKKGYASLEIVQLRDKRLIYLK-QYGGAVKALMGDWYRLRLHKKAGL 221  
 QY 189 -----VLGSMQRLRSMQVNG--SYPRGAKGSRILCTPGWFSQGPDMSCLSRHSI 241  
 Db 222 SLINGINGLIRNPRIQLGKICMLYDIKLDQPLTYNQMWS--GFFPDGSI-----274  
 QY 242 NPYSNRESRLFSFWNDHIEKKRTIIPTEV-----AIKROD 280  
 Db 275 --YLNRSGQIFITAS-----QKNRFLDALVELYGGTYAMVAKOD 313

## RESULT 14

GNLIG5

HIV-1 retropepsin (EC 3.4.23.16) - simian immunodeficiency virus (African green monkey  
 N:Contans: endonuclease (EC 3.1.1.-) retropepsin (EC 3.4.23.16); RNA-directed DNA P  
 C:Species: simian immunodeficiency virus, HIV  
 C:Date: 30-Jun-1989 #sequence\_revision 30-Jun-1989 #text\_change 03-Jun-2002  
 C:Accession: B28873

R:Franchini, G.; Gurgu, C.; Guo, H.G.; Gallo, R.C.; Collalti, E.; Fargnoli, K.A.; Hal  
 Nature 328, 539-543, 1987  
 A>Title: Sequence of simian immunodeficiency virus and its relationship to the human  
 A:Reference number: A28873; MUID:87287229; PMID:3457350

A:Accession: B28873  
 A:Molecule type: DNA  
 A:Residues: 1-1054 <FRA>  
 A:Cross-references: EMBL:M19499

C:Comment: Specific enzymatic cleavages may yield mature proteins including protease,  
 C:Genetics:  
 A:Gene: pol  
 C:Superfamily: pol polyprotein  
 C:Keywords: aspartic proteinase; hydrolase; nucleotidyltransferase; polypeptide; reve  
 F:106-204/Product: retropepsin #status predicted <RTP>  
 F:130/Active site: Asp (shared with dimeric partner) #status predicted

```

Query Match      5.28; Score 92.5; DB 1; Length 1054;
Best Local Similarity 22.88; Pred. No. 18;
Matches 56; Conservative 34; Mismatches 93; Indels 63; Gaps 11;

OY 13 LNSPRKFGVAGRSQCEVLRKGCRLRFLPERGSRCLCYEDGETELTDEYPPVNDNAE----68
DB 296 IIRIPRPAIGLRKRIRITLIDIGDAYFSIP-----LDEFFROYTFTLPSV--NNAEFGKR 347
OY 69 LVILFTLGOAMOG-----VSDIRRFSLAFHE--POVGLIOAQQOLLC--CDEQAPQRRL--118
DB 348 YIKYVLPQGMKSPAFIQTMYTRHVLPEFRKANPDVTLVQYMDLILASDRDLDEHRYVL 407
OY 119 -LADLHNVSQNIARERADPPW-PEGLE-----SRQSSSG 155
DB 408 QLKELNLSIGSTPEKFKQDPFQMMGYLEMTKMLQKLELRQRETWYNQIKRYGV 467
OY 156 LKYSCE-----SRIRYLRVSSYPSIVGAQAEQELRLVLSMCRQLSMQYNG 204
DB 468 LNMAGQIYPIGIRKTHLCRLIRNGKMTLEAVQNTKMAEAEYEENNI-----ILSQDEG 520
OY 205 SYFDNG 210
DB 521 CYIQEG 526

RESULT 15
sex-limited protein Slp(w7) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 27-Oct-1995 #sequence.revision 03-Nov-1995 #text.change 16-Jul-1999
C:Accession: S54784; 148770
R:Ogata, R.T.; Zepf, N.E.
J:Immunol. 147, 2756-2763, 1991
A:Title: The murine Slp gene. Additional evidence that sex-limited protein has no biologic
A:Reference number: S54784; MUID:92013090; PMID:1918990
A:Accession: S54784
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 1-1735 <OGA>
A:Cross-references: EMBL:M64933; NID:9200988; PIRN:AAA0117.1; PTD:9200989
R:Hemenway, C.; Kalf, M.; Stavenhagen, J.; Walthall, D.; Robins, D.
Nucleic Acids Res. 14, 2539-2554, 1986
A:Title: Sequence comparison of alleles of the fourth component of complement (C4) and s
A:Reference number: 148774; MUID:6116748; PMID:3008092
A:Accession: 148770
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 634-641, 'L', 643-828 <RES>
A:Cross-references: EMBL:X06455; NID:954102; PIRN:CAA29761.1; PTD:9899336
C:Genetics:
A:Insertions: 22/2; 86/3; 154/1; 177/3; 207/2; 235/1; 267/2; 302/3; 347/1; 385/3; 444/3; 50
3; 1219/1; 1296/3; 1352/3; 1372/3; 1404/1; 1464/2; 1494/3; 1519/3; 1554/1; 1584/1; 1617/
C:Superfamily: alpha-2-macroglobulin

Query Match      5.18; Score 92; DB 2; Length 1735;
Best Local Similarity 27.68; Pred. No. 38;
Matches 50; Conservative 20; Mismatches 63; Indels 48; Gaps 11;

OY 18 KRGVAGRSQCEV-LRKG--CLRFQLPERSRLCY-----EDGETELTDEYPPVNDNAEL 69
DB 1468 KIGLSGMAIADITLLSGFHALRADLEKITSLSRQYSHRETDGPHVLL-YFDSVPTTRREC 1526
OY 70 VLLTLGOAMOGYVSDIRRFSLAFHEPOVGLIOAQQOLLC-----EQAPQRRL 118
DB 1527 V-----GF-----GASQEVYGLVQPASVNLVDIYSPHKCSVYTAAPTRSQL 1569
OY 119 LADLHNVSQNIARERADPPWPEGLEISRFQSKSGY-LRSCSRIRISYLRVSSPTST 177
DB 1570 LKYL---CSGDVCCACAGKCPRLRLSLRERVEDKQGYRRFAC-----YYHOV-ERGET 1619
OY 178 V 178

DB 1620 V 1620

RESULT 16
sex-limited protein precursor - mouse
N:Alternate names: complement C4-related sex-limited protein
C:Species: Mus musculus (house mouse)
C:Date: 30-Jun-1988 #sequence.revision 30-Jun-1988 #text.change 16-Jul-1999
C:Accession: A29176; B21692; A41195; B29059; B60227; 179467
R:Ogata, R.T.; Seplich, D.S.
J:Immunol. 135, 4239-4244, 1985
A:Title: Murine sex-limited protein: complete cDNA sequence and comparison with murin
A:Reference number: A29176; MUID:86060918; PMID:3840826
A:Accession: A29176
A:Molecule type: mRNA
A:Residues: 1-1736 <OGA>
R:Nonaka, M.; Takahashi, M.; Natsume-Sakal, S.; Nonaka, M.; Tanaka, S.; Shimizu, A.;
Proc. Natl. Acad. Sci. U.S.A. 81, 6822-6826, 1984
A:Title: Isolation of cDNA clones specifying the fourth component of mouse complement
A:Reference number: A94013; MUID:85038607; PMID:6208559
A:Accession: B21692
A:Molecule type: mRNA
A:Residues: 651-749, 'H', 751-774, 'D', 776-802, 921-1040 <NON>
R:Ogata, R.T.; Seplich, D.S.
Proc. Natl. Acad. Sci. U.S.A. 81, 4908-4911, 1984
A:Title: Genes for murine fourth complement component (C4) and sex-limited protein (S
A:Reference number: A41195; MUID:84272739; PMID:6589636
A:Accession: A41195
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1345-1438, 'R', 1440-1544 <OG2>
A:Cross-references: GB:K02293; NID:9199611; PIRN:AAA39662.1; PTD:9199612
R:Toei, M.; Levi-Strauss, M.; Duponchel, C.; Meo, T.
Philos. Trans. R. Soc. Lond. 306, 389-394, 1984
A:Title: Sequence heterogeneity of murine complementary DNA clones related to the C4
A:Reference number: A93753
A:Accession: B29059
A:Molecule type: mRNA
A:Residues: 1255-1335, 'G', 1337-1373 <ROS>
R:Ogata, R.T.; Zepf, N.E.
Eur. J. Immunol. 20, 1607-1610, 1990
A:Title: C4 from C4-high and C4-low mouse strains have identical sequences in the reg
A:Reference number: A60227; MUID:90353398; PMID:2387317
A:Accession: B60227
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1096-1139 <OG3>
A:Cross-references: GB:X55495
R:Nonaka, M.; Kimura, H.; Yeul, Y.D.; Yokoyama, S.; Nakayama, K.; Takahashi, M.
Proc. Natl. Acad. Sci. U.S.A. 83, 7883-7887, 1986
A:Title: Identification of the 5'-flanking regulatory region responsible for the diff
A:Reference number: 159084; MUID:87017050; PMID:3464002
A:Accession: 179467
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-21 <RES>
A:Cross-references: GB:M14226; NID:9199615; PIRN:AAA9684.1; PTD:9554239
C:Genetics:
A:Superfamily: alpha-2-macroglobulin
P:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1736/Product: sex-limited protein #status predicted <MAT>

Query Match      5.18; Score 92; DB 2; Length 1736;
Best Local Similarity 27.68; Pred. No. 38;
Matches 50; Conservative 20; Mismatches 63; Indels 48; Gaps 11;

OY 18 KRGVAGRSQCEV-LRKG--CLRFQLPERSRLCY-----EDGETELTDEYPPVNDNAEL 69
DB 1468 KIGLSGMAIADITLLSGFHALRADLEKITSLSRQYSHRETDGPHVLL-YFDSVPTTRREC 1526
OY 70 VLLTLGOAMOGYVSDIRRFSLAFHEPOVGLIOAQQOLLC-----EQAPQRRL 118

```

```

Db      1527 V-----GF-----GASQEVVYGLVPASAVLYDYXSPDHKCSVFYAAPTRKSQ 1569
OY      119 LADLLINVSQNIATFAEDPPWFEGLESFRQSKSGY-LRYCESIRKIRYLAEVSSYPT 177
Db      1570 LATL-----CSGDVQCAGKCPRLIRSLERREVDKDGYYMRKFC-----YHQV-EYGET 1619
OY      178 V 178
Db      1620 V 1620

RESULT 17
D70441
oxido/reductase iron sulfur protein - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 17-Mar-2000
C:Accession: D70441
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov
e 392, 353-358, 1998
le: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
Reference number: A70300; MUID:98196666; PMID:9537320
Accession: D70441
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-395 <AOE>
A:Cross-references: GB:AE000748; NID:92983960; PIDN:AAC07510.1; PID:g2983963; GB:AE000656
A:Experimental source: strain VFS
C:Genetics:
C:Superfamily: unassigned ferredoxin 2[4Fe-4S]-related proteins; ferredoxin 2[4Fe-4S] hc
F:9-85/Domain: ferredoxin 2[4Fe-4S] homology <FE>

Query Match      5.1%; Score 91.5; DB 2; Length 395;
Best Local Similarity 28.1%; Pred. No. 5.9;
Matches 34; Conservative 19; Mismatches 47; Indels 21; Gaps 7;

OY      229 PRDMSCLSRHSINPYSNRSRIIFSTWNLDHIIEKKR-----TIIPLYEAIKEDGKE 283
Db      206 PDKICCGAPHLT--HNTFAFEKLKEHNKE-LEKYVDALVACPFCGALKEDYGR- 261
OY      284 VDMEYF-YGLIFTSENIKL-----VHIYCHKKTTHKLNCDPSRIYQTRIKRKP 334
Db      262 -DMKVSFEIIANEDIEFGKGEKVYHVHCYTTANKLN--PNNFYALKKVNKAERY 318
OY      335 R 335
Db      319 K 319

CT 18
18
hypotheetical protein - Chlamydia pneumoniae (strain CWL029)
Accession: A72096
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Jun-2000
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606; PMID:10192388
A:Accession: A72096
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-925 <ARN>
A:Cross-references: GB:AE001614; GB:AE001363; NID:94376562; PIDN:AAD18442.1; PID:g437656
A:Experimental source: strain CWL029
C:Superfamily: Chlamydia trachomatis hypotheetical protein CT234

Query Match      5.1%; Score 91.5; DB 2; Length 925;
Best Local Similarity 21.8%; Pred. No. 18;

```

**Matches**    77; Conservative   44; Mismatches   135; Indels    97; Gaps    17;

OY            28 EVLAKGCLRLPGRSGRLCY--EDG-----TELTEPYFSPVDPNAELVLTLLT-GQAMWG 80  
Db            513 ELKRWTSIFPHPAIAAIIHYFAEHDLHTHTHAEDLYDVGDRLLAAITLVRRDAYG 572  
OY            81 YVSDI--RRELSAHEPOVGLIQAAOOILCDEAPQRORLADLLHVNSONIAET---- 134  
Db            573 PYRLADLRKLKEILLNSQPEDIYVGLTIILKEKNRPQNFPLIDPLFNKNDDILIVCKAL 632  
OY            135 ----RAEDPPWFEGLSEHFQ-----SKSGYLRSCE-----SRINSYLEVSSYST-- 177  
Db            633 HTSVANRHKKPCPELLKRLRCOSHNDASOYLTKTISIALDISFPVKOLLMTTSOLKNSTR 692  
OY            178 -----VGAEEQEFRLVYIGSMCORLASMOYNNGSYPRGAAGGSR-----LCPEEG 222  
Db            693 KYAEAMTGELDKVEAAPAFLOVL-----TDEGTHNRCRRIILAANAACKIDN 736  
OY            223 WFSOCGFPMDDSCLSRHSINDPSNRESRIILEFTWNLDHIIEKKRTIIP-----LVEAIR 277  
Db            737 WL-----LKRHAAYKIIVKSKAKALFYGVH-CHYIOKR---YPPTNLSLANTLN 781.  
OY            278 EQDGREVDMEY-FYGLL-----FTSENLEKIIVHIVCHKTKTHKLNCD 317  
Db            782 SNRYAEVNMFLSJGLIGSMESHGSVLLIRALTISKNOK---IKQAQLSELERKNC 831

**RESULT 19**

E81573  
conserved hypothetical protein CP0465 [imported] - Chlamydomophila pneumoniae (strain A  
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 20-Jun-2000  
C:Accession: E81573  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke,  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; Deboy, R.; Kolonay, J.; McClarty, G.; Salzberg  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis Moyn and Chlamydia pneumoniae AR39  
A:Reference number: AB1500; MUID:20150255; PMID:10684935  
A:Accession: E81573  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-925 <RNA>  
A:Cross-references: GB:AE002208; GB:AE002161; NID:g7189387; PIDN:AAF38302.1; PID:g718  
A:Experimental source: strain AR39, HL cells  
C:Genetics:  
A:Gene: CP0465  
C:Superfamily: Chlamydia trachomatis hypothetical protein CT234

**Query Match**                 5.1%; Score 91.5; DB 2; Length 925;  
**Best Local Similarity** 21.8%; Pred. No. 18;  
**Matches**    77; Conservative   44; Mismatches   135; Indels    97; Gaps    17;

OY            28 EVLAKGCLRLPGRSGRLCY--EDG-----TELTEPYFSPVDPNAELVLTLLT-GQAMWG 80  
Db            513 ELKRWTSIFPHPAIAAIIHYFAEHDLHTHTHAEDLYDVGDRLLAAITLVRRDAYG 572  
OY            81 YVSDI--RRELSAHEPOVGLIQAAOOILCDEAPQRORLADLLHVNSONIAET---- 134  
Db            573 PYRLADLRKLKEILLNSQPEDIYVGLTIILKEKNRPQNFPLIDPLFNKNDDILIVCKAL 632  
OY            135 ----RAEDPPWFEGLSEHFQ-----SKSGYLRSCE-----SRINSYLEVSSYST-- 177  
Db            633 HTSVANRHKKPCPELLKRLRCOSHNDASOYLTKTISIALDISFPVKOLLMTTSOLKNSTR 692  
OY            178 -----VGAEEQEFRLVYIGSMCORLASMOYNNGSYPRGAAGGSR-----LCPEEG 222  
Db            693 KYAEAMTGELDKVEAAPAFLOVL-----TDEGTHNRCRRIILAANAACKIDN 736  
OY            223 WFSOCGFPMDDSCLSRHSINDPSNRESRIILEFTWNLDHIIEKKRTIIP-----LVEAIR 277  
Db            737 WL-----LKRHAAYKIIVKSKAKALFYGVH-CHYIOKR---YPPTNLSLANTLN 781  
OY            278 EQDGREVDMEY-FYGLL-----FTSENLEKIIVHIVCHKTKTHKLNCD 317



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 26, 2003, 14:49:22 ; Search time 65 Seconds  
(without alignments)  
1071.445 Million cell updates/sec

Title: US-09-748-451-2

Sequence: 1789  
1 MLQPKSVKLRALRSPKRCG.....SRVYPTRLRKQPVAKRQ 338

Gapop 10.0 , Gapext 0.5

Search table: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Database :

SPREMBL 21:1  
1: sp\_archaea:1  
2: sp\_bacteria:1  
3: sp\_fungi:1  
4: sp\_human:1  
5: sp\_invertebrate:1  
6: sp\_mammal:1  
7: sp\_mhc:1  
8: sp\_organelle:1  
9: sp\_phage:1  
10: sp\_plant:1  
11: sp\_rodent:1  
12: sp\_virus:1  
13: sp\_vertebrate:1  
14: sp\_unclassified:1  
15: sp\_virus:1  
16: sp\_bacteriophage:1  
17: sp\_archaeophages:1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1043	58.3	333	13	Q8UVS5
2	941	52.6	333	13	Q9DGL7
3	419	23.4	80	4	Q96P72
4	419	23.4	110	4	Q96P74
5	419	23.4	115	4	Q96P73
6	321	17.9	450	5	Q9V3H0
7	320	17.9	450	5	Q9V3H0
8	105.5	5.8	571	10	Q48981
9	104.5	5.8	284	11	Q9CZL3
10	104	5.7	1540	6	Q29433
11	102.5	5.6	265	5	Q9U786
12	98.5	5.5	2629	6	Q29021
13	98.5	5.5	684	16	Q8ZG65
14	98.5	5.5	684	16	Q8ZG65
15	98	5.4	594	11	Q8Z787
16	96.5	5.4	271	4	Q60819

17	96.5	5.4	571	16	Q8UKH9	Q8UKH9 agrobacteri
18	95	5.3	383	11	Q88305	Q88305 mus musculu
19	94	5.3	1572	11	Q885N0	Q885N0 mus musculu
20	93.5	5.2	397	8	Q02689	Q02689 podopora a
21	93	5.2	481	5	Q9U787	Q9U787 drosophila
22	92.5	5.2	266	5	Q9V652	Q9V652 drosophila
23	92	5.1	410	4	Q96M12	Q96M12 homo sapien
24	91.5	5.1	395	16	Q67558	Q67558 aquifex neo
25	91.5	5.1	458	13	Q80VR2	Q80VR2 xenopus lae
26	91.5	5.1	925	16	Q9US16	Q9US16 chlamydia p
27	91.5	5.1	925	16	Q9Z8P5	Q9Z8P5 chlamydia p
28	91.5	5.1	1022	15	Q90317	Q90317 chimpanzee
29	91.5	5.1	1022	15	Q90317	Q90317 chimpanzee
30	91	5.1	1719	16	Q9ASB9	Q9ASB9 caulobacter
31	90.5	5.1	265	11	Q9JIT3	Q9JIT3 rattus norv
32	90.5	5.1	340	15	Q8UN04	Q8UN04 chimpanzee
33	90.5	5.1	986	15	Q57059	Q57059 chimpanzee
34	90.5	5.1	1022	15	Q89620	Q89620 chimpanzee
35	90.5	5.1	1056	15	Q04097	Q04097 arabidopsis
36	90	5.0	1044	10	Q9Z0M0	Q9Z0M0 schizosacch
37	90	5.0	1842	3	Q96WT6	Q96WT6 human immun
38	89.5	5.0	337	15	Q9WHZ0	Q9WHZ0 human immun
39	89.5	5.0	340	15	Q8UN03	Q8UN03 chimpanzee
40	89.5	5.0	706	5	Q950A8	Q950A8 caenorhabdi
41	89.5	5.0	1022	15	Q88135	Q88135 chimpanzee
42	89.5	5.0	1022	15	Q87965	Q87965 chimpanzee
43	89.5	5.0	1148	11	Q61372	Q61372 mus musculu
44	89.5	5.0	4116	5	Q96958	Q96958 drosophila
45	89.5	5.0	4120	5	Q9XYD1	Q9XYD1 drosophila

## ALIGNMENTS

RESULT 1	ID	Q8UVS5	PRELIMINARY	PRT	333 AA
AC	Q8UVS5	01-MAR-2002 (TREMBLrel. 20, Created)			
DT	01-MAR-2002 (TREMBLrel. 20, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Caspase-activated Dnase.				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OX	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	Pubmed-11577114;				
RT	Samejima K., Tone S., Earnshaw W.C.;				
RT	Cleavage and Stage I Chromatin Condensation in Apoptosis.;				
RT	J. Biol. Chem. 276:45427-45432(2001).				
DR	EMBL; AF06761; ALJ3716.1; .				
DR	InterPro; IPR003508; CAD.				
DR	Pfam; PF02017; CIDE-N; 1.				
DR	SMART; SM00266; CAD; 1.				
SO	SEQUENCE 333 AA; 98069CD287755E11 CRC64;				
QY	Query Match	58.3%;	Score 1043;	DB 13;	Length 333;
QY	Best Local Similarity	60.1%;	Pred. NO. 7.8e-86;		
QY	Matches 197;	Conservative 51;	Mismatches 74;	Indels 6;	Gaps 2;
Db	6 KSVRLRLRSPKRGVAGRSQEVLRKGLRFLDPERGSRILCYEDGTETLFDYPSVD 65				
Db	6 RGFRLRPPGSAQKFGAAGSLRGLRKCRLQLPLAGSRILCYEDGTETLSAFTPLP 65				
QY	66 NAEVLLTLLGAWGQVYSDIRRFLLSAFEPQVGLIOAAQQLCDEAPAPQRLADLLN 125				
QY	66 QTEVLLRPPSWPGCCGVETRLAALCSRDAVVAEARLRLEDERAPRRRLADLLN 125				

[illegible]

**RESULT 2**

```

ID      09DGJ.7; PRELIMINARY; PRT. 333 AA.
00DGJ.7;
01-MAR-2001 (TREMBLrel. 16, Created)
01-MAR-2001 (TREMBLrel. 16, Last sequence update)
01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE      Caspase-activated Dnase.
05      Brachydanio rerio (zebrafish) (Zebra danio).
06      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
07      Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
08      Cyprinidae; Danio.
09      NCBI_TaxID=7955;
10      [1]
11      SEQUENCE FROM N.A.
12      Liu C.C., Huang C.J.;
13      Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
14      [2]
15      SEQUENCE FROM N.A.
16      Liu C.C., Huang C.J.;
17      "Danio rerio caspase-activated Dnase.";
18      Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
19      EMBL; AF286179; AAF99706.1; -.
20      EMBL; AF426516; AAL40264.1; -.
21      HSSP; 000273; 1IBX.
22      InterPro; IPR003508; CAD.
23      Pfam; PF02017; CIDE-N. 1.
24      SMART; SM00266; CAD. 1.
25      SEQUENCE 333 AA; 38340 MW; 7E15D76E8824F0FB CRC64;

```

Query Match	52.6%	Score 941	DB 13	Length 333
Best Local Similarity	51.5%	Pred. No. 1.2e-76		
Matches 169	Conservative 67	Mismatches 88	Indels 4	Gaps 2

4 KPKSVKLTALISPPKFEVGAAGRCQVLRKGCRLFPOLPEBGSRLCTLEDGTELTDETFPSV 63  
 6 KPKLVKRTISANQARKRYGTALAVNVLKELIKKGCQDLLKAVSSSGVLCVLTEDGTVYEEFFQNL 65  
 64 PDMAELVLTITGQANQGYVSDIRFLSAFHEBQVGLIQAAQULLCDEQAPQRLADLL 123  
 66 PDNFDLVLTLPHGQSNWNGFADEINRYLGDJDRNTEELLVSAQGLTSDERBPKRRRLIGDLL 125  
 QY 124 HNVGQNLIAAETRAEDPPWFEGELESFQSGSLYLRKSCERISLYLREAVSSYPTV- GAEA 182  
 126 SNLRDSELEENLRLOQHDWFEIGIDTFEPKRTSAAMKNCESIRINGLYLEVGYQTQINNAKT 185  
 QY 183 QEEELRVLGVSGMCQRLSRMQYNSGYEDRGAGKGSRLCTPFGMGSCQGPFDMDSCLSRHSIN 242  
 186 KSEKTKVAEALAEKTKAARYNCTYIDRSKQDANRLCTEGEGWFSQCGAFBENSCNFTLHSIN 245  
 QY 243 PYSNRRESILFSTWNLDHIIIEKKRTIIPTLVAIEKQDREYDWEYFGILFTSENIUKY 302  
 246 PYGNRESRILESTWNLDHIEKKRPVLPALAKALEANKSNDINVDYFKILTFREMLIKY 305  
 QY 303 HYGCHKKTTHLNCNDPSRLTYKQYTLK 330

Db 306 HIVCHKGAHELSCDSRKIY---RVKK 330

### RESULT 3

ID	PRELIMINARY;	PRT;	80 AA.
AC Q96P72			
DT 01-DEC-2001 (TrEMBLrel. 19, Created)			
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)			
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)			
DE DNA fragmentation factor B truncated form III.			
DE DFFB			

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Plimates; Catarrhini; Hominae; Homo.  
NCBI\_TaxID=9606;  
OX

RP SEQUENCE FROM N.A.  
RA Bayasas J.R., Yuste V.J., Perera R., Comella J.X.;  
RT "Characterization of splice variants of human caspase-activated DNase  
RT that show CIDE-N structure and function."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AF409062; AAL02007.1; -  
DR InterPro: IPR003508; CAD.  
DR Pfam: PF02017; CIDE-N.1.  
SQ SEQUENCE 80 AA; 9052 MW; 9EB6C9C0B35AE529 CRC64;

```

Query Match          23.4%  Score 419; DB 4; Length 80;
Best Local Similarity 100.0%  Pred. No. 2.1e-30;
Matches      80; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

```

[illegible]

#### RESULT 4

ID	096P74;	PRELIMINARY;	PRT;	110 AA.
AC	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	DNA fragmentation factor B truncated form I.			
GN	DFB.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Carnivora; Hominiina; Homo.			
OX	NCBI_TaxID=9606;			

RP SEQUENCE FROM N.A.  
RA Baygasas J.R., Yuste V.J., Perera R., Comella J.X.;  
RT "Characterization of splice variants of human caspase-activated DNase  
RT that show CIDE-N structure and function."  
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ  
DR EMBL; AF409060; AAL02005.1; -.

DR Pfam; PF02017; CIDE-N; 1.  
SQ SEQUENCE 110 AA; 12106 MW; 34E72888A0F3B909 CRC64;

Query Match	23.48;	Score 419;	DB 4;	Length 110;
Best Local Similarity	100.0%;	Pred. No. 3.3e-30;		
Matches	80;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps

OY	I MLOPKRSVKIRALRSPRKFGVAGSCQEVLRKGLFQLPERGSRICLYEDGETELTDF	60
Dd	I MLOPKRSVKIRALARSPRKFGVAGSCQEVLRRKGCLNFOLPERGSRLCYLTDGTETLDYF	60
Oy	61 PSVPDAAEYLTLTGQAQG 80	

```
DB 61 PSYDPAEVLTLTGQAQMG 80
|||||
RESULT 5
096P73 PRELIMINARY; PRT: 115 AA.
AC 096P73;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE DNA fragmentation factor B truncated form II.
GN DFB.
OS Homo sapiens (Human).
NCBI_TaxID=9606;
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
[1]
SEQUENCE FROM N.A.
RA Bayreuth J.R., Yuste V.J., Perera R., Comella J.X.;
RT "Characterization of splice variants of human caspase-activated DNase
that show CID-N structure and function."
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF409061; AAL02006.1;
DR InterPro: IPR003508; CAD.
DR Pfam: PF02017; CID-N; 1.
SQ SEQUENCE 115 AA; 12849 MW; 4D598FEE32424142 CRC64;

Query Match 23.4%; Score 419; DB 4; Length 115;
Best Local Similarity 100.0%; Pred. No. 3.5e-30;
Matches 80; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MLQPKSVKLRALRSRFRFGVAGRSCEVLRKGCILRFQIPERSRLCYEDGTETLDEYF 60
DB 1 MLQPKSVKLRALRSRFRFGVAGRSCEVLRKGCILRFQIPERSRLCYEDGTETLDEYF 60
OY 61 PSYDPAEVLTLTGQAQMG 80
DB 61 PSYDPAEVLTLTGQAQMG 80
DB 61 PSYDPAEVLTLTGQAQMG 80

RESULT 6
O9V3H0 PRELIMINARY; PRT: 450 AA.
AC O9V3H0;
DT 01-MAR-2000 (TREMblrel. 13, Created)
DT 01-MAR-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
REPA OR DCAD OR CG9414.
REP4 protein (CG9414 protein).
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phlebotomina; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RA MEDLINE-20085845; PubMed-10627165;
RA Itohara N., Nunez G.;
RT "Genes with homology to DFE/CIDDS found in Drosophila melanogaster."
RL Cell Death Differ. 6:823-824 (1999).
[2]
SEQUENCE FROM N.A.
RA Itohara N., Nunez G.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE FROM N.A.
RA MEDLINE-20196006; PubMed-10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Manalidis P.G., Scherf S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang O., Chen L.X.,

BRANDON R.C., ROGERS Y.-H.C., BLAZEK R.G., CHAMPE M., PFELTER B.D.,
WAN K.H., DOYLE C., BAXTER E.G., HELL G., NELSON C.R., MIKLOS G.L.G.,
ABRIL J.F., AGAYANI A., AN H.-J., ANDREWS-PFANKUCH C., BALDWIN D.,
BAILL R.M., BASU A., BAXENDALE J., BAYRAKTAROGULU I., BEASLEY E.M.,
BEESON K.Y., BENOS P.V., BERMAN B.P., BHANDARI D., BOLTSHAKOV S.,
BORKOVA D., BOTCHAN M.R., BOUCH J., BROKSTEIN P., BROTLIER P.,
BURTIS K.C., BUSAM D.A., BUTLER H., DAVENPORT L.B., DAVIES P.,
CHERRY J.M., CAWLEY S., DAHLKE C., DAVENPORT L.B., DAVIES P.,
DE PABLOS B., DELCHER A., DENG Z., MAYS A.D., DEW I., DIETZ S.M.,
DODSON K., DOUP L.E., DOWNES M., DUGAN-ROCHA S., DUNKOV B.C., DUNN P.,
DUBLIN K.J., EVANGELISTA C.C., FERRIZ C., FERREIRA S., FLEISCHMANN W.,
FOSTER C., GABRIELIAN A.E., GARY N.S., GELBART W.M., GLASSER K.,
GLOCKER A., GONG F., GORRELL J.H., GU Z., GUAN P., HARRIS M.,
HARRIS N.L., HARVEY D., HELMAN T.J., HERNANDEZ J.R., HOUCK J.,
HOSTIN D., HOUTSON K.A., HOWLAND T.J., WEI M.-H., IDEYAN C.,
JAJALI M., KALUSH F., KARPEN G.H., KE Z., KENNELSON J.R., KETNUM K.A.,
KIMMEL B.E., KODIRA C.D., KRAFT C., KRAVITZ S., KULP D., LAI Z.,
LAI X., LAU P., LEI Y., LEVITSKY A.A., LI J., LI Z., LIANG Y., LIN X.,
LASKO P., LEE T., MCINTOSH T.C., MCLEOD M.P., MCPHERSON D.,
MERKULOV G., MISHINA N.V., MOKARTY C., MORRIS J., MOSHTELI A.,
MOUNT S.M., MOY M., MURPHY B., MURPHY L., MUZY D.M., NELSON D.L.,
NELSON D.R., NELSON K.A., NIXON K., NUSKERN D.R., PACLEB J.M.,
PALAZZOLO M., PATTMAN G.S., PAN S., POLLARD J., PURI V., REESE M.G.,
REINERT K., REMINGTON K., SAUNDERS R.D.C., SCHEELER F., SHEN H.,
SHUE B.C., SIDEN-KRAMOS I., SIMPSON M., SKUPSKI M.P., SMITH T.,
SPIER E., SPRADLING A.C., STAPLETON M., STRONG R., SUN E.,
SVILSKAS R., TECTOR C., TURNER R., VENTER E., WANG A.H., WANG X.,
WANG Z.-Y., WASSARMAN D.A., WEINSTOCK G.M., WEISENBACH J.,
WILLIAMS S.M., WOODAGE T., WORLEY K.C., WU D., YANG S., YAO Q.A.,
YE J., YEH R.-F., ZAYEVI J.S., ZHAN M., ZHANG G., ZHAO Q., ZHENG L.,
ZHENG X.H., ZHONG F.N., ZHONG W., ZHOU X., ZHU S., ZHU X., SMITH H.O.,
RA GIBBS R.A., MYERS R.W., RUBIN G.M., RUBIN J.C.;
RT "The genome sequence of Drosophila melanogaster."
SCIENCE 287:2185-2195 (2000).
RL EMBL: AF149797; AAF03220.2;
DR EMBL: AEO03639; AAF53276.1;
DR FLYBASE: FBgn0028406; REP4.
DR InterPro: IPR003508; CAD.
DR Pfam: IPR003508; CAD.
DR Pfam: PF02017; CID-N; 1.
DR START: SM00266; CAD: 1.
DR PROSITE: PS00150; CYTOCHROME_C; UNKNOWN.1.
SQ SEQUENCE 450 AA; 52017 MW; CBA693092178BAB CRC64;

Query Match 17.9%; Score 321; DB 5; Length 450;
Best Local Similarity 26.8%; Pred. No. 1.5e-20;
Matches 106; Conservative 61; Mismatches 121; Indels 108; Gaps 17;

OY 17 RKEGVAGRSCEVLRKGCILRFQIPERSRLCYEDGTETLDEYF 75
DB 17 RKEGVAGRSCEVLRKGCILRFQIPERSRLCYEDGTETLDEYF 75
OY 59 RKGIGAGNSLEMLINAKSKPILP--PHYLASDGEVSDDEYLSLP--AQLFIYSG 114
DB 59 RKGIGAGNSLEMLINAKSKPILP--PHYLASDGEVSDDEYLSLP--AQLFIYSG 114
OY 76 QANOGYVSDIRFLSAFHEPQVGLIAQAQQLCD--EQAPOR--QRLADLHNSQNTAA 132
DB 76 QANOGYVSDIRFLSAFHEPQVGLIAQAQQLCD--EQAPOR--QRLADLHNSQNTAA 132
OY 115 P--DAVITWDADFEFEKMKROOSPILKVA--DIYDFIRHQBKRRMITFEHOKORVID 171
DB 115 P--DAVITWDADFEFEKMKROOSPILKVA--DIYDFIRHQBKRRMITFEHOKORVID 171
OY 133 ET-----RAEDPPEEGIESRFGSKSYLTYSCESIRSYLTVESYSYTVGAQAQEEF 186
DB 133 ET-----RAEDPPEEGIESRFGSKSYLTYSCESIRSYLTVESYSYTVGAQAQEEF 186
OY 172 NSKTHLSLKAHEVEMFTGEEFRHSKEAMATRAQIVRGY-----YAKREL 220
DB 172 NSKTHLSLKAHEVEMFTGEEFRHSKEAMATRAQIVRGY-----YAKREL 220
OY 187 LR-----VIGSMCORLREM-----QYNYSQDRGAK----- 212
DB 187 LR-----VIGSMCORLREM-----QYNYSQDRGAK----- 212
OY 221 TRNPLVQNKAQAQVINSVLEKFRILILGCOFFEMEDRCKKHREFLKOHGDEETDAG 280
DB 221 TRNPLVQNKAQAQVINSVLEKFRILILGCOFFEMEDRCKKHREFLKOHGDEETDAG 280
OY 213 -----GGSRLCTPEGFGSCGPF-----DMSCSRISINPYSNR 247
DB 213 -----GGSRLCTPEGFGSCGPF-----DMSCSRISINPYSNR 247
OY 281 RIPSRLKQVIRIYTKENCILDEMSTSLCSLDLGFYCGQSYSRNGSKOKTINPVASR 340
DB 281 RIPSRLKQVIRIYTKENCILDEMSTSLCSLDLGFYCGQSYSRNGSKOKTINPVASR 340
OY 248 ESRILSTWNLDHIEKTKTTIPLVLAIKE-----ODGREVD--WEFYGLL 293
DB 248 ESRILSTWNLDHIEKTKTTIPLVLAIKE-----ODGREVD--WEFYGLL 293
OY 341 ENLLFLVNMLDHQLIELCTLPALVAVNEELVSHPOTKSIHKQVDSIVLEYEFL-I 399
DB 341 ENLLFLVNMLDHQLIELCTLPALVAVNEELVSHPOTKSIHKQVDSIVLEYEFL-I 399
OY 294 FTSENILKLVHIVCHK-----TTRKLNCDPSRIYK 323
```

Db 400 FSLKNTLVHIVCHKAQRNSRNGRLCSDCHEYR 435

## RESULT 7

09NDR2 PRELIMINARY: PRT: 450 AA.  
 ID 09NDR2  
 AC 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE Caspase-activated Dnase.  
 GN REB4 OR DCAD OR CG9414.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]

SEQUENCE FROM N.A.  
 MEDLINE=2023953; PubMed=10777599;  
 Yokoyama H., Mukae N., Sakahira H., Okawa K., Iwamatsu A., Nagata S.;  
 "A novel activation mechanism of caspase-activated Dnase from  
 Drosophila melanogaster";  
 J. Biol. Chem. 275:12978-12986(2000).  
 DR EMBL; AB036773; BAA97120.1; -.  
 DR FlyBase; FBgn0028406; Rep4.  
 DR InterPro; IPR003508; CAD.  
 DR InterPro; IPR00345; CytC\_heme\_bind.  
 DR Pfam; PF02017; CIDE-N; 1.  
 DR SMART; SM00266; CAD; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_1.  
 SQ SEQUENCE 450 AA; 52060 MW; 3D3E4E10AB32E48D CRC64;

Query Match 17.9%; Score 320; DB 5; Length 450;  
 Best Local Similarity 26.8%; Pred. No. 1.8e-20;  
 Matches 106; Conservative 61; Mismatches 121; Indels 108; Gaps 17;

QY 17 RKEGVASRSCOEVLARKCLRFQLPERSRLCYEDGETLED-YFSPVDNAELVLLTG 75  
 DB 59 RKYIGANSLEMLTAKKSKPFLLE-LHLYLASDGEVSDEYLKSLP-AQTLEIVSG 114  
 QY 76 QAMGVYSDIRRFSAFHEPQVGLIOAQQQLCD-EQAPQR-QRLADLLHNSONTAA 132  
 DB 115 P--DAVTTADAEFEKRRQOSPLKVA-DIEYDFIQRPERFRMITEHORARYLD 171  
 QY 133 ET-----RAEDPWFEGLESFQSKGYLRYSCESIRSYLREVSSYSTVGAQAEF 186  
 DB 172 NSKTHLAKAEHVMFTFGEERFHSKEAMATRAQTIVRGY-----YKAKEEL 220  
 QY 187 LR-----VLGSMCQRLKSM--QYNGSYFDGAK----- 212  
 DB 221 TRNPLRYONAKAROVINSVLEKFRYLLIGCDFFSMDFRNCQKHEFLKHLGDEPDAG 280  
 DB 213 -----GGSRLLCPPEGWFCOGPR--DMDSCLSHSTINPSNR 247  
 DB 281 RIPSRLROYIRETKNCLIDEMWSTLCSLDGDFYCGSYSENGNSCSKQHTINPYASR 340  
 QY 248 ESRLFTSWNDHIIEKRTIIPLYEAIKE-----QDGREVD--WEFYGLL 293  
 DB 341 ENLLFQVMNDHQIELCRITLPALVANVELVSHPTQKSHKKQYVDISVLEYFLE-I 399  
 QY 294 FTSNKLVIHIVCHK-----THKINCDSRIYK 323  
 DB 400 FSLKNTLVHIVCHKAQRNSRNGRLCSDCHEYR 435

RESULT 8  
 048981 PRELIMINARY: PRT: 571 AA.  
 ID 048981  
 AC 048981  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE NBS-LRR type resistance protein (Fragment).  
 GN R1.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartioideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=4530;  
 RN [1]

SEQUENCE FROM N.A.  
 RC STRAIN=CV\_NIPONBARE;  
 RX MEDLINE=96081880; PubMed=9419382;  
 RA Leister D., Kurth J., Laurie D.A., Yano M., Sasaki T., Devos K.,  
 RA Graner A., Schulze-Isenhardt P.;  
 RT "Rapid reorganization of resistance gene homologues in cereal  
 genomes";  
 RT Proc. Natl. Acad. Sci. U.S.A. 95:370-375(1998).  
 DR EMBL; AF032688; AAB9695.1; -.  
 DR InterPro; IPR000767; Disease\_resist.  
 DR InterPro; IPR001611; LRR.  
 DR InterPro; IPR002182; NB-ARC.  
 DR Pfam; PF00560; LRR; 3.  
 DR Pfam; PF00931; NB-ARC; 1.  
 DR PRINTS; PR00364; DISEASERISIT.  
 DR NON\_TER 1  
 FT NON\_TER 571  
 FT SEQUENCE 571 AA; 64503 MW; 1288A8734F38B2E6 CRC64;

Query Match 5.9%; Score 105.5; DB 10; Length 571;  
 Best Local Similarity 20.7%; Pred. No. 0.56;  
 Matches 69; Conservative 38; Mismatches 97; Indels 129; Gaps 13;

QY 43 GSRLCYEDGETLEDYFSPV-----PDNAELVLLTQAMGVYSDIR--FLSA 91  
 DB 240 GSLLC-----TKYEDDMKNVLRSEIWELEPDKNNILPALRLSNHLPALIKRFAFCV 294  
 QY 92 FHEPOV-----GLIOAQQQLCDE-----QAPQRLADLL 123  
 DB 295 FHKDYVEKETLVQIMMALGFIQSPGRRTIELSSYFDELGRSFGQHNHGGYVMDAM 354  
 QY 124 HNSONTAAE--TAEADPWFEGLESFQSKGYLRYSCESIRSYLREVSSYSTVGA 181  
 DB 355 HDLAQVSYMDCLYDDPP-----NSSSTSRNSRLSSCHNRKST----- 395  
 QY 182 AQEFLRYLVSQCORLSMOYNGSYFDGAKGSGRLCTPEGWFCOGPRFMDSCLSRHSI 241  
 DB 396 SFEDFL-----GFKKARTLLLLNGYKRSQTSPISD----- 425  
 QY 242 NPYNSRESRLFTSWNDHIIEKRTIIPLYEAIKEODGREVWEYF----- 289  
 DB 426 -----LFLMLRYLHVLLENRRDITELPDSI---GNLKMRLYLNLSGTGTYLPS 471  
 QY 290 -YGLLFTSENKLVHIVCHKRTKHLKNDPSRI 321  
 DB 472 SIGRLFNLIQTLKL-----KNCVLECIPEST 497

RESULT 9  
 09CZL3 PRELIMINARY: PRT: 284 AA.  
 ID 09CZL3  
 AC 09CZL3  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
 DE 11 days embryo cDNA, RIKEN full-length enriched library,  
 DE clone:2700062J12, full insert sequence.  
 GN VIM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 SEQUENCE FROM N.A.

RN	Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX	NCBI_TaxID=9823;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89108006; PubMed-2905687;
RA	Mede N., Ebert D.L., Doers T.M., Newman M., Hasler-Rapacz J.,
RA	Attie A.D., Rapacz J., Smithies O.;
RT	"Molecular genetics of the apolipoprotein B gene in pigs in relation
RL	to atherosclerosis.";
RN	Gene 70:213-229(1988).
RN	[2]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=89108006; PubMed-2905687;
RA	Mede N., Ebert D.L., Doers T.M., Newman M., Hasler-Rapacz J.,
RA	Attie A.D., Rapacz J., Smithies O.;
RT	"Molecular genetics of the apolipoprotein B gene in pigs in relation
RL	to atherosclerosis.";
RN	Gene 69:213-229(1988).
DR	EMBL; M20384; AAA68300.1; -
DR	EMBL; M22647; AAA3097.1; -
KW	Lipoprotein.
FT	NON_TER
SQ	SEQUENCE 1540 AA; 174656 MW; 1E67510A18582F0E CRC64;
Query Match	Best local Similarity 5.8%; Score 104; DB 6; Length 1540;
Matches	66; Conservative 52; Mismatches 123; Indels 98; Gaps 14;
OY	36 RFOLPER-----GSRLC-----LY--EDGETELMEDPFPSPDNAL----- 69
DB	1166 RFLPGRRARYTGDGLCNVMTKEVEGVLSQIYSKINSGLIEILLISFODIMEKSKALK 1225
OY	70 -----VLITFGAMOGYVS DIRRFSAHFHPGYGLIAAOQLTC----- 108
DB	1226 IKFTFDVSYKYLTFFVSEVGEOIKSLSDOVOKALSIDLHSINTETILSELQIFLEGIOEI 1285
OY	109 -DEQAPOKORLADLHNYSONIAETFAEDPPWFEGLESRSOSKSGLYRSCSRI--- 164
DB	1286 EEELIRLEKTKLDLFINDIQONINTFNTPAYAPGFRLKENIDSPFGMLNEFIQNTLMEA 1345
OY	165 RSYLEVESY-----PSNVG-----AEAGEFTLVLSMCGRLSMOYGSGYFDR 209
DB	1346 SOELOLMQYIKAKREYEDPSMGVGMVYKYLEEKYNILINIVDYLK--DFHSKYTVS 1400
OY	210 GAKGSRLCTEGFWFSOQGFPMDSCL-----RHSINPNRESRLTFSTMJLH 260
DB	1404 ATDFASQSSLOVEDFOVE---DIQETSLIADADGKKETKIALESLRKOELL-KSMAN-- 145
OY	261 IIEKRRTIIPLVAIKEODGREVDMEFYGLFTSENTL 299
DB	1458 -----TMKEIISYTHQR----FYKLDFDSDQL 1481
RESULT 11	
ID	Q9U786 PRELIMINARY; PRT; 266 AA.
AC	Q9U786;
DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	DREP3.
GN	REP3 OR DREP3 OR CG8364.
OS	Drosophila melanogaster (fruit fly).
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachytera; Muscomorpha;
OC	Ephyridioidea; Drosophilidae; Drosophila.
OX	NCBI_TaxID=7227;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=2005845; PubMed-10627165;
RA	Inohara N., Nunes G.;
RT	Genes with homology to DF/Crins found in Drosophila melanogaster."
RL	Cell Death Differ. 6:825-824(1999).

DR EMBL; AF149796; AAF03219.1; -  
DR HSP; Q9UHD4; 10AB.  
DR FlyBase; FBgn0028407; Rep3.  
DR InterPro; IPR003508; CAD.  
DR Pfam; PF02017; CIDe-N; 1.  
DR SMART; SMO0266; CAD; 1.  
SO SEQUENCE 266 AA; 29414 MW; 5A9757BCA8AFA645 CRC64;  
Query Match 5.7%; Score 102.5; DB 5; Length 266;  
Best Local Similarity 25.5%; Pred. No. 0.37;  
Matches 42; Conservative 30; Mismatches 72; Indels 21; Gaps 6;  
QY 2 LQKSKYKRLALRPRKFGVAGRGSCQYLRKGCRLPQLPERGSLCLTEDTEL-TEDYF 60  
DB 115 LDNSKPRKIDITRNIRKAVATLSELRTKVSLSKFERAQR-RLHLDCCDEVDDEEYF 172  
QY 61 PSVPDNELVLTGQAMQ--GYVSDIR-FLSAFHEPQVGLIQAAQQLDCBOAPQ 116  
DB 173 STLEPNAELIAVEFGEDQMDSDYNANIRKTSLSA-----QRLSLVSKLQPNYMN 223  
117 RLADLLHNSQNTAETRAEDPWFEGLESFQSKSGYLRYSC 161  
224 DDDDKLSMDPNSLVDITGEPK----DREYSARSDAARLSTE 263  
RESULT 12  
Q29021 PRELIMINARY; PRT; 2629 AA.  
AC Q29021;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
DE Apolipoprotein B (Fragment).  
GN APOB.  
OS Sus scrofa (Pig).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.  
OX NCBI\_TaxID=9623;  
RN [1]  
RX MEDLINE=94014802; PubMed=8409766;  
RA Puteill C., Maeda N., Ebert D.L., Kaiser M., Lund-Katz S.,  
RA Sturley S.L., Kodoyanni V., Grunwald K., Nevlin D.N., Aiello R.J.,  
et al.  
RT "Nucleotide sequence encoding the carboxyl-terminal half of  
RT Apolipoprotein B from spontaneously hypercholesterolemia pigs."  
RL J. Lipid Res. 34:1323-1335(1993).  
DR EMBL; L11235; AAA74655.1; -  
DR InterPro; IPR000510; Oxidized\_nitrogen.  
DR Pfam; PF00148; oxidized\_nitro; 1.  
KW Lipoprotein.  
NON\_TER  
SEQUENCE 2629 AA; 300092 MW; 5389C4C17A87049 CRC64;  
Query Match 5.6%; Score 99.5; DB 6; Length 2629;  
Best Local Similarity 18.7%; Pred. No. 15;  
Matches 72; Conservative 61; Mismatches 140; Indels 113; Gaps 16;  
QY 36 RFQLEPER-----GSRLC-----LY-----EDGTELEDFPSVPMAL----- 69  
DB 2255 RFQLEPGARNTYDGLCNMWTTEGVLSQYISKIHGSLLELLSYFDLAKMSKLNARK 2314  
QY 70 -----VLTGQAMQGYSDIRRLISAFHEPQVGLIQAAQQLC----- 108  
DB 2315 IKFTFDSVKYQYDVEYSEYQSLKSLSDQVOKALSDHSINITEILSELQIFLEGIFQEI 2374  
QY 109 -DEOAPORRLADLLHNSQNTAETRAEDPWFEGLESFQSKSGYLRYSCSRI--- 164  
DB 2375 EEEETRLKEKRLADFDINDIQNTTNTITAPLGFRLKENLDSPEGLNFEIQTILWEA 2434  
QY 165 RSLYREVSY-----PSTVG-----AEOEFLRYLGSQCORLSRQNGSYEDR 209  
DB 2435 SQELQQLHQYIKALRKREYFPDPSMVGWVYKYLEEKYINLIKLVYVK--DFHSKTVS 2492

QY 210 GAKGSLRCLTPGEWFGCGPDMDCIS-----RHSINPSNRESRLFTSWNDH 260  
DB 2493 ATDFASQLSQVEGFVE--DIOEYLSILADADGKREKIAQSSRAQEI--KSMV-- 2546  
QY 261 IIEKRTIIPLVLEI-----KRDGGEVMEYTYGLTFSENKLVIHYCHK--- 308  
DB 2547 -----TWKEIISDYHOQFYIKLQDFSDQSDYVE--KFLAETRLIDLSIQYHM 2594  
QY 309 -----KTHKLNDPSRIYKPTRL 328  
DB 2595 FLRYIMKLKELQSDYVDMRPYIKV 2620  
RESULT 13  
Q82065 PRELIMINARY; PRT; 684 AA.  
AC Q82065;  
DT 01-MAR-2002 (TREMBlrel. 20, Created)  
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
DE DNA helicase IV (EC 3.6.1.-).  
GN HELD OR STM1075.  
OS Salmonella typhimurium.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Salmomella.  
OX NCBI\_TaxID=602;  
RN [1]  
RX MEDLINE=21534948; PubMed=11677609;  
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courtney L., Portwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Waterston R., Wilson R.K.,  
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT Ltr2."  
RL Nature 413:852-856(2001).  
DR EMBL; AE008746; AAL20008.1; -  
DR InterPro; IPR000212; UvrD-helicase.  
DR Pfam; PF00580; UvrD-helicase; 1.  
KW Helicase; Hydrolyase; Complete proteome.  
SO SEQUENCE 684 AA; 78072 MW; A77699E4FCE2989F CRC64;  
Query Match 5.5%; Score 98.5; DB 16; Length 684;  
Best Local Similarity 22.9%; Pred. No. 3.1;  
Matches 89; Conservative 52; Mismatches 148; Indels 99; Gaps 22;  
QY 26 CQEVLRKGCRLPQLPERGSL--CLYEDG-TELEDFPSVPM-----NAEL 69  
DB 152 CREIYMRK-CLAWLQDSEGRQOHQAVADAMLEAHADFFQIESPLNPSQARAVNGES 210  
QY 70 VLTGQAMQGYSDI---RRLISAFHEPQVGLI-----QAQQLDCBOAPQ---Q 116  
DB 211 SLVLVLAGSGSKTVLVARAGWLLARQADAGQITLLAFGRKAEEM--DERIERLATE 268  
QY 117 RLADLLHNSQNTAETRAEDPWFEGLES-----FQSKSGYLRYSCESR---IRS 166  
DB 269 EITARTFHSIALYI--IQGSKKAPVSKLESDARARQDLFRTROOCSEKKAQAKWRO 327  
QY 167 YLREVSYSTVGAEADDEFL--RVLGSQCORLSRQNGSYEDRGAAGKSGSLCTPGWF 224  
DB 328 WLEEMQWVVEGFMFDETLQRLRLARLDRWVSLMRHG-----GAQAEKINAGAP-- 379  
QY 225 SCQGPDMDCSLSHSINPSN---RESRLFTSWNDH-----IIEKRTIIP----- 270  
DB 380 ECRLEFGRIKLPALPKAMSAIKAEANAFS--GLHQAQVTLERGLRSPKHLIYD 437  
QY 271 -----TLVAIKEDGREV-----DMEFYGLFTSENKLVHYI----- 305  
DB 438 EFQDISPQRAALLALAKRQNSQTTLFVAGDQMAIY--RFGAQLDSLTFAPHQTFGGEH 495

```

OY 306 CHKTYTHKLNCDPSRIYKPTRLKRP 333
DB 496 CHDPTTYREN--SRIGDIANRFVOQNP 520

RESULT 14
ID 082787 PRELIMINARY; PRT; 684 AA.
AC 082787;
DT 01-MAR-2002 (Tremblrel. 20; Created)
DT 01-MAR-2002 (Tremblrel. 20; Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21; Last annotation update)
DE Helicase IV (75 kDa helicase).
GN SRY1097.

Salmonella typhi.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
NCBI_TaxID=601;

RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RA MEDLINE=21394947; PubMed=11677608;
RA Parhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moulie S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrett B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18."
RL Nature 413:848-852(2001).
DR EMBL; AL627269; CAD08201.1;
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KW Helicase; Complete proteome.
SQ SEQUENCE 684 AA; 78038 MW; 5E9077CD302A198E CRC64;

Query Match 5.5%; Score 98.5; DB 16; Length 684;
Best Local Similarity 22.9%; Pred. No. 3.1; Mismatches 148; Indels 99; Gaps 22;
Matches 89; Conservative 52;

OY 26 COBYLRKGCIRFOLPERGSL--CLYEDG-TELTEDYPPSPVD-----NAEL 69
DB 152 CREIMWK-CLAWLADSESSRQHNQAYADMLAHADFTQIESPLNPQARAVNGES 210
DB 70 VLLTIGQAMQGYSDI--RRFISAFHEPQVGLI-----QAAQOLLCDEQAPOR--Q 116
DB 211 SLVYLAGAGSGKTSVYVARAGMILLARQADAGQILLARGRKAAEEM--DRIRERLHTE 268
OY 117 RLADLHNVSQNTIAETRAEDPPWEGLES---RFQSKSGYLRYSCESR-----IRS 166
DB 269 EITATRFHSLALYI-IOGSKKAPVYKLSDDATANHQLFLHTWRQCSKKQAQKGRQ 327
OY 167 YLRVSSYPSTVGAEAOEFL--RVLSMCQRLSMQVSYFDDGAGKGSRLCPREGWF 224
DB 328 WLEEEKQWVYVPEGNWDETLQRLRLADRWVSLRMHG---GAQAEMAGAPE--- 379
OY 225 SCGGPDMQSCLSRSINIPYSN--RESRLTFSTNMLD---IIEKKRTIIP----- 270
DB 380 EGRLEFGKRIKLMVAILKAWKSALKANAVDF--GLIHQAVILLEKGRFISPKHILYD 437
OY 271 -----TLVEYIKRODGRV-----DMREYFGLLFTSENKLYHIV----- 305
DB 438 EFQDISPQAAILLERKQNSQTLTFLAVGDQWQALY--RFSGAQLSTTAAHQTFEGEGE 495
OY 306 CHKTYTHKLNCDPSRIYKPTRLKRP 333
DB 496 CHDPTTYREN--SRIGDIANRFVOQNP 520

RESULT 15
OY 062238 PRELIMINARY; PRT; 594 AA.
AC 062238;
DT 01-NOV-1996 (Tremblrel. 01; Created)
DT 01-NOV-1996 (Tremblrel. 01; Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20; Last annotation update)
DE Sex-muscle protein S1P(w7) alpha-gamma chain (Fragment).
OS Mus musculus (Mouse).
OC Mammalia; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Eukaryota; Euteleostomi; Rodentia; Sclerogamathi; Muridae; Murinae; Mus.
GN NCBI_TaxID=10090;

RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2(C4)(H)SLP(A);
RA MEDLINE=6176748; PubMed=3008092;
RA Hemmaway C., Kalit W., Steyenhagen J., Walthall D., Robins D.;
RT "Sequence comparison of alleles of the fourth component of complement
RT (C4) and sex-limited protein (S1P)."
RL Nucleic Acids Res. 14:2539-2554(1986).
DR EMBL; X06454; CAA29760.1;
DR InterPro; IPR001599; Macroglobin2.
DR Pfam; PF00207; A2M; 1.
DR Pfam; PF01759; NTR; 1.
FT NON_TER 1 303
FT CHAIN 304 594
FT CHAIN 304 594
SQ SEQUENCE 594 AA; 66454 MW; 9465C569DA37768 CRC64;

Query Match 5.5%; Score 98; DB 11; Length 594;
Best Local Similarity 24.1%; Pred. No. 2.8; Mismatches 82; Indels 72; Gaps 12;
Matches 58; Conservative 29;

OY 18 KFGVAGSGCEV-LRGG-CLRFOLPERGSLCY-----EDTELTEDYPPSPVDNML 69
DB 327 KLIGSGMALADITLISGFALRGDEKLTSLDRYVSHFPTDQPHYL-YFDSVYTRREC 385
OY 70 VLLTIGQAMQGYSDIRFLSFFHPQVGLQAAQOLCD-----EQARQRL 118
DB 386 V-----GF-----GASQEVYGLVQSSNAVLDIYSPDKSCVFIAATKSQL 428
OY 119 LADLHNVSQNTIAETRAEDPPWEGLESRFQSKSGY-LRYSC----- 160
DB 429 LATL--GSDGVCQCAQGCPLRLSLRRVEDKGRMRACVYPRVEYGTVAFLRED 485
OY 161 -----ESRISYLRVSSYPSTVGAEAOEFLRVLSGMCQRLSMQVSYFDDGAG 212
DB 486 GRAAFRLFSKITQVLFHFTDPMASIG---QTRNPLSRTSCLRL-LEPNKEYLIMGMD 539
OY 213 G 213
DB 540 G 540

RESULT 16
OY 060819 PRELIMINARY; PRT; 271 AA.
AC 060819;
DT 01-AUG-1998 (Tremblrel. 07; Created)
DT 01-AUG-1998 (Tremblrel. 07; Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20; Last annotation update)
DE D1934G17.1.2 (Chloride channel protein CLC-6b) (Isoform 2)
DE (Fragment).
GN CLC6..
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euteleostomi; Primates; Catarrhini; Hominoidea; Homo.
GN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC STRAIN=H.;
RA Erlingsson H.;
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021155; CAA15952.1;

```

DR InterPro: IPR001807; Cl-channel\_volt.

DR Pfam: PF00654; voltage\_CIC.1.

DR PRINTS: PR00762; ClCHANNEL.

FT NON\_TER 1

SQ SEQUENCE 271 AA; 30375 MW; 4D9374B5BFIABEBA CRC64;

Query Match 5.4%; Score 96.5; DB 4; Length 271;

Best Local Similarity 23.6%; Pred. No. 1.3; Mismatches 71; Indels 91; Gaps 14;

Matches 61; Conservative 35; Mismatches 71; Indels 91; Gaps 14;

18 KFGVAGSCQEVLRKGLRQLEPERGSLCLYEDGETELTEDEYFSPVDNAELVLTLCQA 77

60 KFGVAGSCQEVLRKGLRQLEPERGSLCLYEDGETELTEDEYFSPVDNAELVLTLCQA 106

78 WQGVSDIRRFSLAFHEBPVGLIOAAQQLCDQAPORRLADLHNVSONIAETRAE 137

107 GSG-IPEVKCYLNGVKVP--GIVR-LRTLLCK-----VLGVLEFVAGGL----- 146

138 DPEWFEGLSFRQSGYLRYSCSRIRSYLREVSYPVYGAQEAEFLAVLSMCORL 197

147 -----FEKEGPMHS-----GSVYGA-GLPQFSI-----SL 173

198 RSMQYNGSYF--DRGAKGSRIC--TPGWFSC-----OGPFDMDSCLSRHS 240

174 RKIOFNFPYFRSDRYGKRQERLCISRGSCSCSFRGANRGYLVQSRGCVL----- 225

241 INPYNSRESRLTFSTNMU 258

226 LEPRAHVESALLFHVCHL 243

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

DR EMBL: AE008936; AAL45813.1; -

DR EMBL: AE007884; AAK90497.1; -

DR EMBL: Complete proteome.

SQ SEQUENCE 571 AA; 62030 MW; 92F4894EAA5934F7 CRC64;

Query Match 5.4%; Score 96.5; DB 16; Length 571;

Best Local Similarity 21.6%; Pred. No. 3.6; Mismatches 122; Indels 115; Gaps 20;

Matches 79; Conservative 49; Mismatches 122; Indels 115; Gaps 20;

10 LRALSPKPFVAGSCQEVLRKGLRQLEPERGSLCLYEDGETELTEDEYFSPVDNAEL 69

259 LASLVGPKSL--GRHVEIYREADIVLITFRN-----QNGTDNMQYSP-----DAOI 306

70 VLTLLGQAWQGVSDIRRFSLAFHEBPVGLIOAAQQLCDQAPORRL-LADI-LHNV 127

307 IHIDTPDQEVGRNREAIRLVDAEETLAGL-----RERIRCDLHLRTVS 351

128 ONIAETPAEDPPEEGLESFRQSGYLRYSCSRIRSYLREVSYPVYGAQEAEFLAVLSMCORL 181

352 REAVCNRLRES--WKFNDDR--RGY--YSSEASPLRPERIMAELOGVIDENTVYVAD 403

182 AQEEFLRYLSM--CORLRSMQYNGSYFDGAKG-----GSLRCTPE----- 221

404 ASYSMMVYLGGLRLISSERTVLV-----PGIAGLGWGVPLAIGAKYAPESDVVALYG 457

222 -GWFSCGPFDMDSCLSRHSINPYNSRESRLTFSTW-NLDHIIERKRTIPTLVEAKR-- 277

458 DGCFV-----HS-----MALEETLY--RMQIPITITIVLNG 486

278 ----EODGREVDWEYFGLTSENKLVHLYCHKTKTHKLNCDPSRIYKQTLKRRP 333

487 ILGFORDAEIVKFKYTTACHFAE-----VDHVKIAHACGCAVRYTOPDLKAHMK 539

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB

QY

DB



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 26, 2003, 15:23:03 ; Search time 297 Seconds  
(without alignments)

2562.880 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1.789  
Sequence: 1 MLOKPKSVKLRLALRSPKRGK.....SRVYQTRLRKQPVAKRQ 338

ng table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
Xgapop 10.0	Ygapext 0.5	
Ygapop 6.0	Ygapext 7.0	
Delop 6.0	Delext 7.0	

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

```

-MODEL=frame_plus_p2n.model -DEV=x1h
-O=/gen2_1/USPTO.spool/US09748451.unat.21052003.153829.16273/apd_query.fasta.1.519
-DB=N_Geneseq.101002 -OFMT=faststep -SUFFIX=ring -MIMMARCH=0.1 -LOOPEL=0
-LOOPEXT=0 -UNIT=bits -STAR=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR.SCORE=0 -THR.MAX=100 -THR.MIN=0 -ALIGN=20
-MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09748451.ecgn.1.1.263.funat.21052003.153829.16273 -NCPV=6 -ICPV=3
-NO.MAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database:

1: N\_Geneseq.101002.DAT:  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:  
25: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1789	100.0	AAH74642	Nucleotide sequenc
2	1789	100.0	AAH02498	Human DNA fragment
3	1786	99.8	AAH19907	Caspase activated
4	1763	98.5	AAH38992	Human DNA fragment
5	1416	79.2	AAH21173	Mouse caspase-acti
6	1416	79.2	AAH21180	Mouse caspase acti
7	372	20.8	AAH18941	human poly(ADP-ribo
8	321	17.9	ABH15179	Drosophila melanog
9	303.5	17.0	ABH15179	Human gene express
10	295	16.5	ABH15178	Drosophila melanog
11	124.5	7.0	ABH44763	Human CDNA differe
12	106.5	6.0	AAH21607	Neisseria meningit
13	106.5	6.0	AAH21607	N. meningitidis B
14	105	5.9	AAH1540	Human DNA sequence
15	102.5	5.7	AAH09307	Human Vimentin cod
16	102.5	5.7	AAH09307	Human DNA sequence
17	101.5	5.7	AAH04122	Porcine acylglucos
18	101	5.6	ABH51606	Mouse ischaemic co
19	101	5.6	ABH14002	Drosophila melanog
20	98	5.5	ABH14002	Human polynucleoti
21	97.5	5.4	ABH44206	CDNA #149 encoding
22	97.5	5.4	ABH44206	Human glutamate re
23	97.5	5.4	ABH44206	Human glutamate re
24	96	5.4	ABH44206	Human glutamate re
25	96	5.4	ABH44206	Human glutamate re
26	96	5.4	ABH44206	Human glutamate re
27	96	5.4	ABH44206	Human glutamate re
28	95.5	5.3	ABH07309	Human nervous syst
29	95.5	5.3	ABH07309	Human nervous syst
30	95.5	5.3	ABH07309	Human nervous syst
31	95.5	5.3	ABH07309	Human nervous syst
32	94.5	5.3	ABH07309	Human nervous syst
33	94.5	5.3	ABH07309	Human nervous syst
34	94.5	5.3	ABH07309	Human nervous syst
35	93.5	5.2	AAH51968	Human polynucleoti
36	93	5.2	AAH51968	Human polynucleoti
37	92.5	5.2	AAH51968	Human polynucleoti
38	92.5	5.2	AAH51968	Human polynucleoti
39	92.5	5.2	AAH51968	Human polynucleoti
40	92.5	5.2	AAH51968	Human polynucleoti
41	92.5	5.2	AAH51968	Human polynucleoti
42	92.5	5.2	AAH51968	Human polynucleoti
43	92.5	5.2	AAH51968	Human polynucleoti
44	92.5	5.2	AAH51968	Human polynucleoti
45	92.5	5.2	AAH51968	Human polynucleoti

## ALIGNMENTS

RESULT 1	
AAH74642	AAH74642 standard; cDNA: 2839 BP.
ID	AAH74642
AC	AAH74642
XX	15-OCT-2001 (first entry)
DT	XX
DE	Nucleotide sequence of human DNA fragmentation factor 40 (DFP40).
XX	XX
DE	Human; DNA fragmentation factor; DFP40; DFP45; apoptosis; DNase;
XX	XX
KW	molecular chaperone; cancer cell; ss.
XX	XX
OS	Homo sapiens.
XX	XX
FT	Location/Qualifiers
FT	132..1148
FT	CD5
FT	/*tag - a

/product= "DNA fragmentation factor 40 (DFP40)"

FT  
 XX  
 PN US2001011078-A1.  
 XX  
 PD 02-AUG-2001.  
 XX  
 PE 22-DEC-2000; 2000US-0748451.  
 XX  
 PR 16-APR-1998; 98US-0061702.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Wang X, Liu X;  
 XX  
 DR WPI: 2001-496169/54.  
 DR P-PSDB: AAG63592.  
 XX  
 PT New DNA fragmentation factor polypeptides and polynucleotides, useful  
 PT for inhibiting the growth of cancer cells, as well as for inducing  
 PT apoptosis of cells -

Claim 18; Page 44-45; 56pp; English.

The present sequence encodes a human DNA fragmentation factor subunit  
 of 40 kDa, designated DFP40. The specification also describes DFP45.  
 CC DFP40 is capable of inducing apoptosis, and may contain a nuclear  
 CC localization fragment. DFP45 acts as a molecular chaperone to direct  
 CC the folding of DFP40. Although all DFP40 is associated with  
 CC DFP40, DFP activity only occurs once DFP40 is complexed with DFP45.  
 CC The DFP polypeptides and polynucleotides are useful for inhibiting  
 CC the growth of cancer cells, and for inducing apoptosis of cells.

XX  
 SQ Sequence 2839 BP; 644 A; 701 C; 755 G; 739 T; 0 other;

## Alignment Scores:

Pred. No.: 1-78e-176 Length: 2839  
 Score: 1789.00 Matches: 338  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Caps: 0

US-09-748-451-2 (1-338) x AAH74642 (1-2839)

QY 1 MetLeuGlnIysProIysSerValIysLeuArgAlaLeuArgSerProAlaGlyPheGly 20  
 DB 132 ATGCTCCAGAAAGCCCAAGAGCGCTGAGCTGGCGCCGACCCGAGGAAAGTGGCG 191  
 QY 21 ValAlaGlyArgSerGlyGlnIleValIleuArgIysGlyCysLeuArgPheGlnLeuPro 40  
 DB 192 GTGGCTGGCCGAGAGCTGCGAGAGGCTGCGCAAGGGCTGTCTCCGCTCCAGCTCCCT 251  
 41 GlnArgIysSerArgLeuGlyLeuArgIysGlyGlnIleuArgIysGlyGlnIleuArgIys 60  
 252 GAGGCGGGTTCGGGCTGTGCTGTACGAGATGGACGAGGCTGACGAGGAAATATTCTTC 311  
 61 ProSerValProAspAsnAlaGlnIleuValIleuLeuThrLeuGlyGlnAlaTrpGlnGly 80  
 312 CCCAGTCTCCGCAACAGCCGAGCTGTGCTGCTCACTTGGGCGCAAGGCTGCGAGGCG 371  
 QY 81 TyrValSerAspIleArgArgPheLeuSerIleAlaPheHisGluProGlnValAlaGlyLeu 100  
 DB 372 TATGTGAGCGACATCAGGCGCTTCTCTGATTCATTTTCACGAGCCAGGTGGGCTCATTC 431  
 QY 101 GlnAlaIleGlnIleuLeuGlyAspGlnIleuAlaProGlnIleuArgGlnIleuAla 120  
 DB 432 CAGGCGCCGACGAGCTGT 491  
 QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaIleuArgAlaGlnIleuArgProPro 140  
 DB 492 GACCTCTGCAACAGCTGAGGCAAGACATCGCGGCGAGACCCGGGCTGAGAGACCCGCG 551  
 QY 141 TrpPheGlnIleuGlnIleuArgIysSerGlyTrpLeuArgTrpSerGly 160

DB 552 TGGTTGAAGCTTGGAGTCCCGATTTTCAGAGCAAGTCGTGCTATCTGAGATACGCTGT 611  
 QY 161 GluSerArgIleArgSerTrpLeuArgIleValSerSerTrpProSerThrValAla 180  
 DB 612 GAGAGCCGGATCCGAGATTACCTGAGGAGGTGAGTCTTACCCCTCCACAGTGGGTGCG 671  
 QY 181 GlnAlaGlnIleuGlnIleuValIleuArgIysSerMetCysGlnIleuArgSerMet 200  
 DB 672 GAGCTCAGAGAAATTCCTCGGGCTCTCGCTTCATGTCCAGAGAGCTCCGGTCCATG 731  
 QY 201 GlnTrpAsnGlySerTrpPheAspArgIleAlaIysGlySerArgLeuCysThrPro 220  
 DB 732 CAGTACATGGCAGCTACTTGCACAGAGAGACCAAGGGCGGACCCGCTCTGCACACG 791  
 QY 221 GluGlyTrpPheSerCysGlnIleuProPheAspMetAspSerCysLeuSerArgHisSer 240  
 DB 792 GAGGCTGTGCTCTCTGCGAGGTCCTTTACATGAGACGCTCTTATCAAGACATCC 851  
 QY 241 IleAsnProTrpSerAsnArgGluSerArgIleLeuPheSerThrTrpAsnLeuAspHis 260  
 DB 852 ATCAACCCCTACAGTACAGAGGAGAGAGAGATCTTACAGACCTGGAACTGGATCAC 911  
 QY 261 IleIleGlnIysIysArgTrpIleIleProThrIleuValGlnAlaIleIysGlnIleu 280  
 DB 912 ATAAATGAAAGAAAGACGACCATCATCTCTGAGGAGCAATTAAGAAACAAGAT 971  
 QY 281 GlnArgGlnValAspTrpGlnTrpPheTrpGlyLeuLeuPheThrSerGlnAsnLeuIys 300  
 DB 972 GGAAGAGAGAGGAGCTGGAGATATTTATGCGCTCTTTTACTCAGAGAACCTTAAA 1031  
 QY 301 LeuValHisIleValCysHisIysIysThrThrHisIysLeuAsnCysAspProSerArg 320  
 DB 1032 CTAGTGCACATTTGCTGCCATTAAGAAACCAACCAAGCTCACTGTGAGCCGAGCAGA 1091  
 QY 321 IleTrpIysProGlnThrArgLeuArgIysGlnIleuValArgIysGlnIleu 338  
 DB 1092 ATCTACAAACCCAGACAGAGGTGAAGCGGAGAGGCTGTGCGGAAAGCCGAG 1145  
 RESULT 2  
 AAD02498  
 ID AAD02498 standard; cDNA: 2839 BP.  
 AC AAD02498;  
 XX  
 DT 24-APR-2001 (first entry)  
 DE Human DNA fragmentation factor 40 (DFP40) cDNA.  
 KW Human; DNA fragmentation factor; DFP; apoptosis; molecular chaperone;  
 KW gene therapy; hyperproliferative disorder; therapy; tumour; restenosis;  
 KW psoriasis; angiogenesis; cancer; cytosolic; neoplasia; ss.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 132..1148  
 FT /\*tag= a  
 FT /product= "DNA fragmentation factor 40 (DFP40)".  
 XX  
 RN US6165737-A.  
 XX  
 PD 26-DEC-2000.  
 XX  
 PF 16-APR-1998; 98US-0061702.  
 XX  
 PR 16-APR-1998; 98US-0061702.  
 XX  
 PA (TEXA ) UNIV TEXAS SYSTEM.  
 XX  
 PI Wang X, Liu X;  
 XX  
 DR WPI: 2001-090481/10.

DR P-PSDB: AAY72467.  
 XX Identifying modulator of human DNA fragmentation factor 40, for  
 PR treating cancer, involves contacting cell or cell-free composition  
 PR comprising DFF40 with candidate substance and comparing apoptosis with  
 PR control  
 XX  
 PS Example 1: Column 77-80; 52pp; English.  
 XX  
 CC The present sequence is a human DNA fragmentation factor 40 (DFF40)  
 CC CDNA capable of inducing apoptosis. DFF40 acts as a molecular chaperone  
 CC comprising 40kDa and 45kDa subunits. DFF40 acts as an inhibitor  
 CC to facilitate the appropriate folding of DFF40 and acts as an inhibitor  
 CC for DFF40. DFF40 and DFF45 are used in gene therapy. The modulators of  
 CC human DFF40 activity are useful for inducing apoptosis and for treating  
 CC hyperproliferative disorders such as restenosis, psoriasis, metastatic  
 CC tumours, angiogenesis and benign and malignant neoplasms. They are also  
 CC used for treating cancers of the brain (glioblastoma, astrocytoma,  
 CC oligodendroglioma and ependymoma), lung, liver, spleen, kidney, lymph  
 CC node, pancreas, small intestine, blood cells, colon, stomach, breast,  
 CC endometrium, prostate, testicle, ovary, skin, head and neck, oesophagus,  
 CC bone marrow, blood, other tissue and multi-drug resistant cancer.  
 CC  
 SQ Sequence 2839 BP; 644 A; 701 C; 755 G; 739 T; 0 other:  
 Alignment Scores:  
 Pred. No.: 1.78e-176 Length: 2839  
 Score: 1789.00 Matches: 338  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 22 Gaps: 0  
 US-09-748-451-2 (1-338) x AAD02498 (1-2839)  
 OY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20  
 DB 132 ATGCTCCAAAGACCCCAAGGCGTGAAGCTGCGGCCCTGCGAGCCGAGAGAGTTCGGC 191  
 OY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40  
 DB 192 GTGGCTGGCGGAGGCTGCGAGAGGCTGCGAGAGGCTGCTCCGCTTCACAGCTCCCT 251  
 OY 41 GluArgLysArgLysLeuArgLysGluArgLysGluArgLysGluArgLysGluArgLys 60  
 DB 252 GAGCGCGGTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311  
 OY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGly 80  
 DB 312 CCAGGTGTTCCGAGAACGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371  
 OY 81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIle 100  
 DB 372 TATGTGAGGACATCAGAGCGCTCTCCTCAGTGCATTTCAGAGCCACAGGTGGGCTCATTC 431  
 OY 101 GlnAlaAlaGlnGluLeuLeuCysAspGlnGlnAlaProGlnArgGlnArgLeuAla 120  
 DB 432 CAGGCCGCCACAGACGCTGCTGTGATGAGAGAGGCCCCACAGAGGAGAGAGCTGCTGCT 491  
 OY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrArgAlaLysProPro 140  
 DB 492 GACCTCTGACAGAGTACGACAGAACATCGCGGCCGAGACCCGGCTGAGGACCCGCGG 551  
 OY 141 TrpPheGlnGlyLeuGlnSerArgPheGlnSerArgLysSerGlyLysLeuArgLysCys 160  
 DB 552 TGGTTTGAAGGCTGGAGTCCGATTCAGAGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCT 611  
 OY 161 GluSerArgLysArgSerLysLeuArgLysValSerSerLysProSerLysArgLysAla 180  
 DB 612 GAGAGCCGATCCGAGGATACCTGAGGAGGTGAGTGCCTCAACCCCTCCACAGTGGGTGG 671  
 OY 181 GluAlaGlnGluThrPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMet 200

DB 672 GAGGCTCAGAGAGAAATTCCTGCGGCTCCTCCATGTCACAGAGGCTCCGCTCATG 731  
 OY 201 GlnTyrAsnGlySerLysPheAspArgGlyAlaLysGlyLysSerArgLeuCysThrPro 220  
 DB 732 CAGTCAATGACACTCTCTTCAGACAGAGAGGCGGAGCCGCTCTCTCCACACCG 791  
 OY 221 GluGlyTrpPheSerCysGlnGlyProPheAspMetLysSerCysLeuSerArgHisSer 240  
 DB 792 GAGGCTGTTCTCTCTCCAGGCTCCCTTTGACATGAGACACTCTTATACACACTCC 851  
 OY 241 IleAsnProTyrSerAsnArgLysSerArgLysIleLeuPheSerThrTyrPheAsnLeuAspHis 260  
 DB 852 ATCAACCCCTACAGTAACAGAGAGAGAGAGATCTCTTACACACTGGAACCTGATCAC 911  
 OY 261 IleIleGlyLysLysArgThrIleIleProThrLeuValGluAlaIleLysGluGlnAsp 280  
 DB 912 ATATATGAAAGAAAGAAACGACCATTCATTCCTACACTGTGGAAAGCAATTAAGAAACAGAT 971  
 OY 281 GlyArgGlyLysLysArgThrLysLysLysLysLysLysLysLysLysLysLysLysLys 300  
 DB 972 GGAAGAGAGTGGAGTGGAGTATTTTATGAGCTGCTCTTTTACCTCAGAGAACCTAAAA 1031  
 OY 301 LeuValHisIleValCysHisLysLysLysLysLysLysLysLysLysLysLysLysLys 320  
 DB 1032 CTAGTGCACATTTGTGTGCTATAGAAACACCCACACACTGATGTGACCCGAGCAG 1091  
 OY 321 IleTyrLysProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338  
 DB 1092 ATCTACAAACCCCAAGAGTTGAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1145  
 DB AAXI9907 standard; DNA; 1017 BP.  
 AAXI9907;  
 11-JUN-1999 (first entry)  
 DE Caspase activated nuclease CPAN encoding DNA.  
 XX  
 KW Caspase activated nuclease; CPAN; apoptosis; human; heart attack;  
 KW stroke; tumour; ss.  
 OS Homo sapiens.  
 PN W09910501-A1.  
 XX  
 PD 04-MAR-1999.  
 PF 19-AUG-1998; 98MO-US17214.  
 XX  
 PR 06-JAN-1998; 98US-0072192.  
 PR 22-AUG-1997; 97US-0056904.  
 PR 22-AUG-1997; 97US-0056907.  
 XX  
 PA (CHIR ) CHIRON CORP.  
 PI Halenbeck R, Macdonald H;  
 XX  
 DR WPI: 1999-190620/16.  
 DR P-PSDB: AAY04123.  
 XX  
 PR New isolated caspase activated nuclease - used to develop products  
 PR which can inhibit or activate apoptosis for use in treating e.g.  
 PR heart attack, stroke or tumours  
 XX  
 PS Claim 1; Page 38; 61pp; English.  
 CC The present sequence encodes a caspase activated nuclease (CPAN). Active  
 CC CPAN proteins can be used for identifying compounds which inhibit  
 CC apoptosis. Inhibitors of apoptosis can be used therapeutically, e.g.  
 CC where cells are damaged and the apoptotic pathway is initiated such as  
 CC by heart attack or stroke. The inactive CPAN proteins can be used for

CC Identifying compounds which activate apoptosis. Compounds which activate  
CC apoptosis can be used to treat tumours. The association of CPAN with a  
CC caspase-sensitive inhibitor suggests that when cells enter into  
CC apoptosis, the activation of caspases will trigger CPAN to become  
CC active.

XX Sequence 1017 BP: 236 A; 293 C; 303 G; 185 T; 0 other:

# Alignment Scores:

Aligned. No.: 8,34e-177 Length: 1017  
Score: 1786.00 Matches: 337  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.70% Mismatches: 0  
Query Match: 99.83% Indels: 0  
DB: 20 Gaps: 0

US-09-748-451-2 (1-338) x AAX19907 (1-1017)

```

QY 1 MetLeuGlnIlyProLysSerValIlyLeuAAlaLeuArgSerProAlqIlyPheGly 20
    1 ATGCTCCAGAGAGCCCAAGAGCGTGAAGCTCGGCGCTGGCAGCCGAGAGAAATTCGGC 60
    21 ValAlaGlyArgSerCysGlnGluValLeuArgIlyCysLeuAArgPheGlnLeuPro 40
    61 GTGGCTGGCCGAGACTCCAGAGAGTCTGCGCAGAGGCTGTCTCCCTCCAGCTCCCT 120
    QY 41 GluArgGlySerArgLeuCysLeuArgIlyAspGlyThrGluLeuThrGluAspTyrPhe 60
    121 GAGCGCGGTTCGCGGCTGTGCTGTACAGAGATGGACGAGCGAGCGAAGATTAATCTTC 180
    QY 61 ProSerValProAspAsnAlaGluLeuLeuLeuThrLeuGlnAlaIleArgGlnGly 80
    181 CCCAGGTCTCCGCAACAGCCGAGCTGCTGCTCCTCAGCTTGGGCGAGGCTGGCAGGCG 240
    QY 81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuLe 100
    241 TATGTAGGCAACATCAGAGCGCTTCTCAGTCAATTCACAGAGCAAGAGTGGGCTATC 300
    DB 101 GlnAlaAlaGlnLeuLeuLeuCysAspGlnAlaProGlnArgGlnArgLeuLeuAla 120
    301 CAGGCGCCCGCAGCAGCTGCTGTGTGATGACAGCCGACAGAGCGAGGCTGTGGCT 360
    QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrArgAlaGluAspProPro 140
    361 GACCTCTGCAACAGCTCAGCAGCAGACATCGCGGCGCAGAGCCGCGGTGAGAGCCGCG 420
    QY 141 TrpPheGlnGlyLeuGlnLeuArgPheGlnSerIlySerIlyTyrLeuAArgTyrSerCys 160
    421 TGGTTTGAAGGCTTGGAGTCCCGATTTCAGAGCAAGCTGGCTATCTGAGATACAGCTGT 480
    DB 161 GluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAla 180
    481 GAGAGCGGAGTCCGAGACTTACCTGAGGAGGTGAGCTCTTACCCCTCCAGCGTGGGTGG 540
    QY 181 GlnAlaGlnGlnGluPheLeuArgValLeuGlySerMetCysGlnAlaGluLeuArgSerMet 200
    541 GAGGCTCAGAGAGAAATTCCTGCGGCTCTGCGGCTCCCAATGTCAGAGAGCTCCGGTCAATG 600
    DB 201 GlnTyrAsnGlySerTyrPheAspArgGlyAlaIlyGlyLeuSerArgLeuCysTyrPro 220
    601 CAGTACAAATGAGCACTACTCTGACAGAGAGCAAGGCGGCGGCGCTCTGACACCG 660
    QY 221 GluGlyTyrPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240
    661 GAAGGCGGTTCCTGCTCCAGGCGCTTGTACATGACAGCTCTTATGAAAGACATCC 720
    DB 241 IleAsnProTyrSerAsnArgGlnSerArgIleLeuPheSerThrTrpAsnLeuAspHis 260
    721 ATCAACCCCTTACAGTACAGGAGAGAGATCTCTTCCAGACCTGGAACCTGAGATCAC 780
    QY 261 IleIleGlnIlyLysArgThrIleIleProThrLeuValGlnAlaIleIlyGlnGlnAsp 280
    781 ATATATGAAAAGAAAGACGACCATCATCTCTTCACTGTGTGAGAGCAATTAAGAAACAAGAT 840
  
```

```

QY 281 GlyArgGluValAspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGlnAsnLeuLys 300
    841 GGAAGAGACAGTGGAGCTGGAGTATTTTANGGCTGCTTTTACCCTCAGAGAACTTAAA 900
    DB 301 LeuValHisIleValAlCysHisLysLysTyrThrHisLysLeuAsnCysAspProSerArg 320
    901 CTAGTGCACATGTGCTCCATTAAGAAAACCAACCCACCAAGCTCAACTGTGACCCGAGCAGA 960
    QY 321 IleTyrLysProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338
    961 ATCTACAAACCCGACAGACAGGTTAAGCGAGAGAGCTGTGCGGAAACGCCAG 1014
    RESULT 4
    ID AA238992 standard; cDNA; 2839 BP.
    AC AA238992;
    XX 22-FEB-2000 (first entry)
    DE Human DNA fragmentation factor DFF40 nucleotide sequence.
    XX Human: DNA fragmentation factor; DFF40; DFF45; apoptosis; gene therapy;
    XX Human: cytosolic; growth; tumour; ss.
    OS Homo sapiens.
    XX MO954482-AL.
    XX 28-OCT-1999.
    XX 16-APR-1998; 98WO-US07895.
    XX 16-APR-1998; 98WO-US07895.
    XX 16-APR-1998; 98WO-US07895.
    XX (TFXA ) UNIV TEXAS SYSTEM.
    XX Wang X, Liu X;
    XX MPI: 2000-052702/04.
    XX P-PSDB: AAY57440.
    PT DNA fragmentation factor DFF40 involved in apoptosis and related
    PT polynucleotide
    PS Claim 18; Page 125-127; 154pp; English.
    CC The present sequence encodes a human DNA fragmentation factor,
    CC designated DFF40. Also described are: (1) a method of inducing apoptosis
    CC in a cell comprising providing the cell with DFF40 which results in
    CC apoptosis; (2) a method for inhibiting the growth of a cancer cell
    CC comprising contacting a cancer cell with a DNA fragmentation factor
    CC designated DFF40 under conditions permitting the uptake of the DNA
    CC fragmentation factor by the cell where the presence of the DFF40 into
    CC the cell induces apoptosis; (3) a method for treating cancer comprising:
    CC (a) encoding a DFF40 DNA fragmentation factor; and (b) a promoter active
    CC in the tumour cell, where the promoter is operably linked to the region
    CC encoding the DNA fragmentation factor, under conditions permitting the
    CC uptake of the nucleic acid by the tumour cell; (4) a method of
    CC identifying a modulator of DFF40; and (5) a method of producing a
    CC functional DNA fragmentation factor. An expression construct encoding a
    CC DNA fragmentation factor DFF40 and DFF45 complex is provided to a cell to
    CC induce apoptosis, especially in tumour cells. DFF40 is used to inhibit
    CC the growth of a cancer cell, especially in humans.
    XX
    SQ Sequence 2839 BP: 642 A; 699 C; 750 G; 737 T; 11 other:
  
```

# Alignment Scores:

Aligned. No.: 9.26e-174 Length: 2839  
Score: 1763.00 Matches: 334  
Percent Similarity: 98.82% Conservative: 0  
Best Local Similarity: 98.82% Mismatches: 4

```
Query Match: 98.55% Indels: 0
DB: 21 Gaps: 0
US-09-748-451-2 (1-338) x AA238992 (1-2839)

QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20
DB 132 ATGCTCCAGAACCCCAAGAGCTGCGGGCCCTCCGACCCCAANNNNNNNNNN 191
QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40
DB 192 GTGGCTGGCCGGAGCTCCAGAGAGGTGGCGGAAGGGGTGTCTCCGCTTCCACTCCCT 251
QY 41 GluArgGlySerArgLeuCysLeuArgLysPheGlyPheGlyPheGlyPheGlyPhe 60
DB 252 GAGCGCGGTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311
QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
DB 312 CCCAGTGTCCCGACACAGCCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 371
QY 81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIle 100
DB 372 TAGGTGAGCGACATGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 431
QY 101 GlnAlaAlaGlnGlnLeuLeuCysAspGlnAlaProGlnAlaArgGlnAlaLeuAla 120
DB 432 CAGGCGCCCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 491
QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaAlaGluThrArgAlaGluAspProPro 140
DB 492 GACCTCTGACAAACGTCAGCAGAACATCGCGCGAGAACCCGGGCTGAGAGACCCGCG 551
QY 141 TyrPheGlnGlyLeuGlnSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCys 160
DB 552 TGGTTGAAGCTTGAGATCCCGATTTCAAGACAATGCTGCTGCTGCTGCTGCTGCTGCT 611
QY 161 GluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAla 180
DB 612 GAGAGCGGATCCGGAGTACTGAGAGGAGTGAAGTCTTACCCTCCACAGTGGCGCG 671
QY 181 GluAlaGlnGlnGluPheLeuArgValLeuGlySerMetCysGlnAlaLeuArgSerMet 200
DB 672 GAGCTCAGAGAGATTCCTGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
QY 201 GlnTyrAsnGlySerTyrPheAspArgGlyAlaLysGlyGlySerArgLeuCysThrPro 220
DB 732 CAGTCAATGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 791
QY 221 GluGlyTyrPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240
DB 792 GAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 851
QY 241 IleAsnProTyrSerAsnArgGlnSerArgIleLeuPheSerThrTyrPheAsnLeuAspHis 260
DB 852 ATCAACCCCTACAGTACAGAGAGAGAGAGATCTCTTCAGACACTGGAMCCTGATCAC 911
QY 261 IleIleGluLysLysArgThrIleIleProThrLeuValGluAlaIleLysGluGlnAsp 280
DB 912 ATATATAGAAAGAAAGAACGACCATCATCTCCACTGCTGGAAGCAATTAAGAAAGAAAT 971
QY 281 GlyArgGluValAspTyrGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLys 300
DB 972 GGAAGAGAAAGTGAAGTGAAGTATTTTATGCGCTGCTTTTACTCAGAAACCTTAA 1031
QY 301 LeuValHisIleValCysHisLysLysThrHisLysLeuAsnCysAspProSerArg 320
DB 1032 CTATGTCACATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1091
QY 321 IleTyrLysProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338
DB 1092 ATCTCAAAACCCAGACAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1145
```

```
RESULT 5
AA221173
ID AA221173 standard; cDNA; 1038 BP.
AC AA221173;
DT 23-NOV-1999 (first entry)
DE Mouse caspase-activating Dnae encoding cDNA.
KW Mouse; caspase activating Dnae; CAD; Inhibitor; ICAD; ICAD-L; ICAD-S;
XX cell death; ss.
OS Mus sp.
XX
XX Key Location/Qualifiers
XX CDS 1..1035
XX FT /tag="a" "CAD"
XX FT /product="CAD"
XX FT /note="caspase-activating Dnae"
XX
XX JP11239495-A.
XX
XX PD 07-SEP-1999
XX
XX 25-DEC-1998; 98JP-0369222.
XX
XX 25-DEC-1997; 97JP-0369356.
XX
XX (OSAB-) 2H OSAKA BIOSCIENCE KENKYUSHO.
XX
XX WPI: 1999-554024/47.
XX
XX P-PSDB; AAY29931.
XX
XX Inhibitor for caspase-activating Dnae - useful for preventing cell
XX death and is useful as reagent for cell death
XX
XX Example: Page 15-16; 26pp; Japanese.
XX
XX The present invention describes an inhibitor (ICAD) for caspase-
XX activating Dnae (CAD). Two ICAD's are specifically claimed: a long
XX chain form (ICAD-L) and a short chain form (ICAD-S). ICAD and ICAD cDNA
XX can be used in the prevention of cell death, and can be used as a
XX research reagent for cell death. The present sequence encodes mouse
XX CAD which is used in the exemplification of the present invention.
XX
XX Sequence 1038 BP; 236 A; 293 C; 307 G; 202 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3,886-138 Length: 1038
XX Score: 1416.00 Matches: 261
XX Percent Similarity: 88.17% Conservative: 37
XX Best Local Similarity: 77.22% Mismatches: 38
XX Query Match: 79.15% Indels: 2
XX DB: 20 Gaps: 1
XX
XX US-09-748-451-2 (1-338) x AA221173 (1-1038)
QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20
DB 10 GTGCTCGGCAACCCCAAGAGCTGCGGGCCCTCCGACCCCAANNNNNNNNNN 69
QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40
DB 70 GTGGCTGGCCGGAGCTCCAGAGAGGTGGCGGAAGGGGTGTCTCCGCTTCCACTCCCT 129
QY 41 GluArgGlySerArgLeuCysLeuArgLysPheGlyPheGlyPheGlyPheGlyPhe 60
DB 130 ATGCGCGGTCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189
QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuLeuLeuLeuLeuLeuLeuLeu 80
DB 190 CCGGCGCTTCCAGACAGAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249
```

[illegible]

XX JP11239494-A.  
XX  
XX 07-SEP-1999.  
XX  
XX 25-DEC-1998; 98JP-0369093.  
XX  
XX 25-DEC-1997; 97JP-0369443.  
XX  
XX (OSAB-) ZH OSAKA BIOSCIENCE KENKYUSHO.  
XX  
XX WPI: 1999-554023/47.  
XX  
XX P-PSDB; AAT9936.  
XX  
XX New caspase-activating DNase - useful for eliminating cancer cells  
XX  
XX Claim 4; Page 14; 20pp; Japanese.  
XX  
XX The present sequence encodes mouse caspase activating DNase (CAD).  
XX  
XX CAD can be used for the removal of cancer cells.  
XX  
XX  
XX  
XX Sequence 1038 BP; 236 A; 293 C; 307 G; 202 T; 0 other;  
XX  
XX  
XX Alignment Scores:  
XX  
XX Pred. NO.: 3.88e-138 Length: 1038  
XX  
XX Score: 1416.00 Matches: 261  
XX  
XX Percent Similarity: 88.17% Conservative: 37  
XX  
XX Best Local Similarity: 77.22% Mismatches: 38  
XX  
XX Query Match: 79.15% Indels: 2  
XX  
XX DB: 20 Gaps: 1  
XX  
XX  
XX US-09-748-451-2 (1-338) x AA22180 (1-1038)  
XX  
XX  
XX QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20  
XX ::::::::::: ||||| ::::::::::: |||  
XX Db 10 GTGCTCGCCGCAACCAATGCTGCAAGTTGGAGCCCTACATAGGCGCTGCAGTTCGGC 69  
XX  
XX QY 21 ValAlaGlyAraSerGlySerGlnValLeuArgLysGlyLysLeuArgPheGlnLeuPro 40  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 70 GTGGCGGCGCCGAGCTGCCAGAGCTGCTGAAGGGCTCGTCCTCCACCTCCG 129  
XX  
XX QY 41 GluArgGlySerArgLeuGlySerLeuTyrgLnaAspGlyThrGluLeuThrGluAspTyrPhe 60  
XX ||||| ::::::::::: ||||| ::::::::::: |||  
XX Db 130 ATGCCGGGTTCCGGCTGTGCTGTACGAAATGGCAGGAGTGGAGCGAGCATGCTTC 189  
XX  
XX QY 61 ProSerValProAspAlaGlnLeuLeuValLeuLeuThrLeuGlyGlnAlaTyrGlnGly 80  
XX ||| ::::::::::: ||||| ::::::::::: |||  
XX Db 190 CCGGGGCTTCCCAACGACGCTAGCTCCATTGCTCACCGGTGGGAGACCTGGCATGGC 249  
XX  
XX QY 81 TyrValSerAspIleArgIleArgPheLeuSerAlaPheHisGlnProGlnValGlyLeuIle 100  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 250 TATGTGAGTACATCAACAGCTTTCCTCAAGTGTGTTAATGAGCCACATGCGCGGTATC 308  
XX  
XX QY 101 GlnAlaAlaGlnLeuLeuLeuGlyAspGlnAlaAlaProGlnArgGlnArgLeuAla 120  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 310 CAGGCTGCACGCAACATGCTGTCAAGATGAGCGCCACATGAGGCAAAAGCTCTGGCC 369  
XX  
XX QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaGlnIuThrArgAlaGluAspProPro 140  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 370 GATCTTCTGTATCACTGTGAGCCAGAAATATTACTGAGAGACCCGGAGACAGACCATCC 429  
XX  
XX QY 141 TrpPheGlnGlyLeuLysSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCys 160  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 430 TGGTTTGAAAGTTTGGAGTCAAGATTCACGAATTAAGTGGGCTATCTGACATACAGCTGT 488  
XX  
XX QY 161 GluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAla 180  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 490 GAGAGTCGAGTCCGGGGTACTTAAGAGAGGTGAGCGCTTACACOTCTATGTGTGAGAA 549  
XX  
XX QY 181 GluAlaGlnGlnLysPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMet 200  
XX ||||| ::::::::::: ||||| ::::::::::: |||||  
XX Db 550 GCAGCTCAAGAAAGTACCTGGAGATCTCTGGCTCCATGTCACGAAAGCTCAATCGGTG 609

```

OY 201 GlnTyrAsnGlySerTyrPheAspArgGlyAlaIleGlySerArgLeuCythrPro 220
DB 610 CAGTACAAAGGAGGAGTATTCGACAGAGGAGGAGCCAGAGCCGCTGTGACATCA 669
OY 221 GluGlyTyrPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240
DB 670 GAAGATGATGTTCTCTCGACAGGAGCCCTTGACCTGGAAGCTGTCTTCCAACTCC 729
OY 241 IleAsnProTyrSerAsnArgGluSerArgIleLeuPheSerThrTyrAsnLeuAspHis 260
DB 730 ATCAACCCCTATGCAACAGAGAGAGCCGATCTCTTCACTGGAACCTGATCAT 769
OY 261 IleIleGlyLysArgThrIleIleProThrLeuValGluAlaIleLysGluGlnAsp 280
DB 790 ATATATAGAAAGAGAGCCAGCCGTGACCCAGCTGCGTGAAGCCATC-----CAGGAT 843
OY 281 GlyArgGluValAspTyrGlnTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLys 300
DB 844 GGGAGGAGGTGAAGTGGAGTACTTTCACGCTCTCTTCACTGCGGAGAACCTGAG 903
OY 301 LeuValHisIleValCysHisLysLysThrHisLysLeuAsnGlyAspProSerArg 320
DB 904 CTGGTCACATCGCTGCGCACAGAGAGCCACACACAGCTGAGTGGAGCCGAGTAGG 963
OY 321 IleTyrLysProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338
DB 964 ACTATCGGCTTCACAGACAGATCCAGAGAGAGAGCCGCTGTGGAAGAGAGCC 1017
RESULT 7
AA189481/c
ID AA189481 standard; cDNA; 379 BP.
AC AA189481;
AC 06-NOV-2001 (first entry)
DE Human polynucleotide SEQ ID NO 9541.
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
OS Homo sapiens.
PN WO200164835-A2.
OY 07-SEP-2001.
DE 26-FEB-2001; 2001WO-US04927.
XX 28-FEB-2000; 2000US-0515126.
PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT;
PI WPI; 2001-514838/56.
DR P-PSDB; AAO09550.
XX The invention relates to human polynucleotides (AA189941-AA193841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX

```

```

CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.
XX
SQ Sequence 379 BP; 88 A; 90 C; 97 G; 104 T; 0 other;

Alignment Scores:
Pred. No.: 1,07e-29 Length: 379
Score: 372.00 Matches: 76
Percent Similarity: 75.24% Conservative: 3
Best Local Similarity: 72.38% Mismatches: 10
Query Match: 20.79% Indels: 16
DB: 22 Gaps: 2

US-09-748-451-2 (1-338) x AA189481 (1-379).
OY 234 SerCysLeuSerArgHisSerIleAsnProTyrSerAsnArgGluSerArgIleLeuPhe 253
DB 341 TCTGCTCTGAGC-----CTCATG 324
OY 254 SerThrTyrAsnLeuAspHisIleIleGlyLysArgThrIleIleProThrLeuVal 273
DB 323 AGTCGCTGGGCGC-----TACAGATGAAAAAGAACGCCACATCTGTCAGATGCTG 270
OY 274 GluAlaIleLysGluGlnAspGlyValArgGluValAspTyrGlnTyrPheTyrGlyLeuLeu 293
DB 269 GAAGCAATTAGAGAGAGATGAAAGAGAGAGTGGAGTGGAGATTTTATGCGCTGCTT 210
OY 294 PheThrSerGluAsnLeuLysSerValHisIleValCysHisLysLysThrHisLys 313
DB 209 TTACCTCAGAGAGACTTAACCTAGTGTGACATCTGTCCATTAAGAAAAACCCACAG 150
OY 314 LeuAsnCysAspProSerArgIleTyrLysProGlnThrArgLeuLysArgLysGlnPro 333
DB 149 CTCACCTGTGACCAAGACAGATATCTACATTCGCCAGCAAGCTTGAAGCGCAGAGGCT 90
OY 334 ValArgLysArgGln 338
DB 89 GTGGCGAAGAGACAG 75

RESULT 8
ABL15179
ID ABL15179 standard; cDNA; 1353 BP.
AC ABL15179;
AC 26-MAR-2002 (first entry)
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 40019.
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ss.
XX Drosophila melanogaster.
PN WO200171042-A2.
OY 27-SEP-2001.
XX 23-MAR-2001; 2001WO-US09231.
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX (PEKE ) PE CORP NY.
PA Venter JC, Adams M, Li PMD, Myers EW;
PI
XX

```

DR WPI: 2001-656860/75.  
DR P-SDB: ABB71076.

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -

PS Claim 1: SEQ ID NO 40019; 21bp + Sequence Listing: English.

XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
CC sequences (AB101840-AB16175) and the encoded proteins  
CC (AB857737-AB872072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

Sequence 1353 BP; 379 A; 337 C; 340 G; 297 T; 0 other;

Alignment Scores:

pred. No.: 1.42e-23 Length: 1353  
Score: 321.00 Matches: 106  
Percent Similarity: 42.17% Conservative: 61  
Best Local Similarity: 26.77% Mismatches: 121  
Query Match: 17.94% Indels: 108  
DB: 23 Gaps: 17

US-09-748-451-2 (1-338) x AB15179 (1-1353)

QY 17 ATGATGpGhegYValAlaGlySerCysGlnGluValLeuArgLysGlyCysLeuArg 36  
DB 175 AGGAATAACGAAATCGAGCAATCCCTGGAATGCTGATAGGCAAGCAAAAGCAAG 234  
QY 37 PheGlnLeuProGluArgGlySerArgLeuGlyCysLeuArgGlyValGlnLeuThr 56  
DB 235 TTCCCGCTTCGCG-----CCACATTTTGTATTTGGCTGCTGATGCGCTGAGGTGCC 288  
QY 57 GlnuSP---TyrPheProSerValProAspAsnAlaGluLeuValLeuThrLeuGly 75  
DB 289 GACGATGATGATCTAAAGGCTTACC-----GCCAGACGCTGTCATATGTTCTGCG 342  
QY 76 GlnAlaTrpGlnGlyTyrValSerAspIleArgArgPheLeuSerAlaPheHisGluPro 95  
DB 343 CCG-----GATGACGATTCACAAACAGATCCGATTTTGATTGAGAAATGCGACAA 396  
QY 96 GlnValGlyLeuIleGlnAlaAlaGlnLeuLeuGlyCysAsp-----GluGlnAlaPro 113  
DB 397 CAATCGCTTGTGCTAAAGGTGGCT---GACATTTTCTACGATTTTATGACACACATCCG 453  
QY 114 GlnArg---GlnArgLeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaAla 132  
DB 454 GAATAGTTCCGCGCATGATACGAGATGACAAACAAAGACGACGCTCTGGAT 513  
QY 133 GlnuThr-----ArgAlaGluAspProProTyrPheGlnGlyLeuGlu 146  
DB 514 AACGACAAAGCCACCTCAGCTTAAGCGACAGCTGAGTGTTCACGCGGCGAG 573  
QY 147 SerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerArgIleArgSer 166  
DB 574 GAGGCTTCCACCTTAAGAGAGAGATGCGCCACAGCTGCCACACGTCGCGAGC 633  
QY 167 TyrLeuArgGluValSerSerTyrProSerThrValGlnAlaGlnGlnGluPhe 186  
DB 634 TACTAC-----TATTAAGGCCAAGAGACAGCTG 660  
QY 187 LeuArg-----ValLeuGlySerMetCys 194  
DB 661 ACCGCGAATCCCTTGTACCGCCAAATGCCAAGGCCGACAAAGATGATTAACCTGTGCTG 720

QY 195 GlnArgLeuArgSerMet-----GlnTyrAsnGlySerTyrPheAspArgGly 210  
DB 721 GAGAAATTCGATACCTGCTCATCGCTTGCACATTTTCTCCATGATGATTTGACCGGAT 780  
QY 211 AlArg----- 212  
DB 781 TGCAAGCAAAAGCATGATTCCTGAGCAGCAGCTGGCGAGAAACAGACGCTGCC 840  
QY 212 ----- 212  
DB 841 AGCATACCCACAGACACTGAGCAGTGTATAGGAGTACACCAAGAAACTGATC 900  
QY 213 -----GlySerArgLeuGlyCysThrProGlnGlyTyrPheSerCysGlnGlyPro 229  
DB 901 CTCGACGAATGCTCAGCTCTCTTGTGCTCCAGCTGGTGAATCTTATGCGACGGCTCC 960  
QY 230 Phe-----AspMetAspSerCysLeuSerArgTyrHisSerIleAsnProTyrSerAsnArg 247  
DB 961 TATTCGAGAAATGCAACAGCTGCTTAAAGCAGCAGCACTATTAATCCGTACGCTTCCGT 1020  
QY 248 GluSerArgIleLeuPheSerThrTyrAsnLeuAspHisIleIleGlnLysLysArgThr 267  
DB 1021 GAGATTCATTTTGTCTTTCAGTGTGATGTGACACCAACCAATGCAATCTGCCGACCA 1080  
QY 268 IleIleProThrLeuValGlnAlaIleLysGlu----- 278  
DB 1081 ATCTTCCTCCGCTGTTGCAATGTGGAAGAACTGTGAGTCAATCCGACACCAAGTGT 1140  
QY 279 -----GlnAspArgGluValAsp-----TyrGluTyrPheTyrGlyLeuLeu 293  
DB 1141 TCAATTCATAAGAAAGCAAGTGGTCAATATCTCACTGAGTCACTGCTTCTAGAA--ATA 1197  
QY 294 PheThrSerGluAsnLeuValHisIleValCysHisLysLys----- 309  
DB 1198 TTTCCTCCAAAGATTCCTCAACTGACATGTGTCTGATGAAAGCCGACGCTCA 1257  
QY 310 -----ThrThrHisLysLeuAsnCysAspProSerArgIleTyrLys 323  
DB 1258 AACCGTTCGAATGGTGGCTGCTGCTGCCACTGTGATGATATCCG 1305

RESULT 9  
AAZ16100  
ID AAZ16100 standard; cDNA: 735 BP.  
AC AAZ16100;  
AC AAZ16100;  
DT 12-Oct-1999 (first entry)  
DE Human gene expression product cDNA sequence SFO ID NO:3570.  
XX  
XX Human; gene; gene expression product; diagnosis; therapy; probe;  
KW detection; mapping; tissue typing; profiling; forensic; cancer;  
KW genetic analysis; colorectal cancer; breast cancer; lung cancer; ss.  
XX Homo sapiens.  
OS  
XX  
XX W0938972-A2.  
PN  
XX  
XX 05-AUG-1999.  
PD  
XX  
XX 28-JAN-1999; 99WO-US01619.  
PF  
XX  
XX 03-APR-1998; 98US-0080666.  
PR 28-JAN-1998; 98US-0072910.  
PR 24-FEB-1998; 98US-0075954.  
PR 31-MAR-1998; 98US-0080114.  
PR 03-APR-1998; 98US-0080515.  
PR  
XX  
XX (CHIR ) CHIRON CORP.  
PA (HYSE-) HYSEQ INC.  
PA  
XX  
XX Cirvenjakov R, Dickson M, Dirmanc R, Dirmanc S,  
PI Escobedo J, Garcia PD, Garcia V, Gleese K, Innis MA;

I Jones WT, Kassam A, Kennedy GC, Kita D, Labat I;  
 PI Iamson G, Ieshkowitz D, Pot D, Randazzo P, Reinhard C;  
 PI Strache-Crain B, Sudduth-Klinger J, Williams LT;  
 DR WPI: 1999-494092/41.  
 XX  
 XX Novel human genes and their expression products which are  
 PT differentially expressed in different cell types  
 PS  
 PS Claim 1: Page 1704: 2479pp; English.  
 CC The present invention describes a library of human polynucleotides  
 CC comprising the sequences given in AA12532 to AA17779. Also described is  
 CC a method of detecting differentially expressed genes correlated with the  
 CC cancerous state of a mammalian cell, comprising detecting at least one  
 CC differentially expressed gene product in a test sample from a cell  
 CC suspected of being cancerous, where the gene product is encoded by one  
 CC of the 5248 polynucleotide sequences given in AA12532 to AA17779. The  
 CC polynucleotides can be used as a source of primers and probes, which can  
 CC be used for a variety of purpose, e.g. detection of expression levels,  
 CC mapping, tissue typing or profiling, forensics, genetic analysis and  
 CC detection of polymorphisms. Polypeptides encoded by the polynucleotides  
 CC can be used for raising antibodies for experimental, diagnostic and  
 CC therapeutic purposes. The polynucleotides may also be used to construct  
 CC arrays for diagnostics (which may be used to determine function of an  
 CC encoded protein) to detect differences in expression levels between  
 CC two cells (e.g. to identify abnormal or diseased tissue in a human, to  
 CC identify a genetic predisposition or susceptibility to a disease such as  
 CC cancer). The polynucleotides of the invention are especially used in the  
 CC diagnosis, prognosis and management of colorectal cancer, breast cancer,  
 CC and lung cancer. The polynucleotides can also be used to screen for  
 CC peptide analogues and antagonists.  
 XX  
 XX Sequence 735 BP: 141 A; 189 C; 190 G; 177 T; 38 other:  
 XX  
 Alignment Scores:  
 Pred. No.: 3,98e-22 Length: 735  
 Score: 303.50 Matches: 79  
 Percent Similarity: 68.64% Conservative: 32  
 Best Local Similarity: 66.95% Mismatches: 9  
 Query Match: 16.96% Indels: 1  
 DB: 20 Gaps: 1  
 US-09-748-451-2 (1-338) x AA16100 (1-735)  
 QY 82 ValSerAspIleArgGatGPhleuSerAlaPheHisGluProGlnValGlyLeuIleGln 101  
 |||||  
 336 GTGAGCAGACATGAGCGCGCTTCTCAGTGNATTTCCAGACCCACAGCTGGGCTTATCCAA 395  
 |||||  
 102 AlaAlaGlnGlnLeuLeuCyAspGluGlnAlaProGlnArgGlnArgLeuAlaAsp 121  
 |||||  
 396 GCCGNCACACACCTGCTGTGTGATGACAGACGCCACAGAGNANAACTGCTGGCTGAC 455  
 |||||  
 122 LeuLeuHisAsnValSerGlnAsnIleAlaIleGluThrArgAlaGluAspProTyr 141  
 |||||  
 456 CTCCTGCACACAGCTCANCATCATGCGGAGACAGACNNCNGTATGAGA -CCCCGTGG 514  
 |||||  
 142 PhgGluGlyLeuGluSerArgPheGlnSerIleSerGlyTyrLeuArgTyrSerCysGlu 161  
 |||||  
 515 NTTGAAGC-TTGGATTTTCGATTTCCANAGCANTNCGCTATCTCANTANCTGTTNAG 573  
 |||||  
 162 SerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGlu 181  
 |||||  
 574 AGCGGATCCCGACATCTAGCTAGAGAGATGAGCTCTCA -CCATTCACGATGGAGCGAG 632  
 |||||  
 QY 182 AlaGlnGluGluPhe-----LeuArgValLeuGlySerMetCys 194  
 |||||  
 DB 633 NCTAA-CAGGAATTCGCGGTCTTCTCATATGACAGAGCTCCGTCAATCATCAGC 685  
 |||||  
 RESULT 10  
 ABL15178/c  
 ID ABL15178 standard; cDNA: 3758 BP.

QY 86 ArgArgPheLeuSerAlaPheHisGlu-----ProGlnValGlyLeuIle 100  
 DB 2002 TGTCTCTATTGATTCGCGATTTTGTGAGTGTGAGAGATGCGACCAATCCCTTTCTA 1943  
 QY 101 GlnAlaIaGlnGlnLeuLeuCysAsp-----GlnGlnAlaProGlnArg---GlnArg 117  
 DB 1942 AAGGTGGCT---GACATTTTCTACGATTTTATTCAGAACAGATCCGGAAATTTCCGGCGC 1886  
 QY 118 LeuLeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaIaGluThr----- 134  
 DB 1885 ATGATTAACGCGAGTATGAAACACCAAAAGACGAGCGCTCGATTAACAGACAGACCCAC 1826  
 QY 135 -----ArgAlaGlnAspProPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 151  
 DB 1825 CTCAGCTTAAAGCGCGAGACACGTGGAGCTTCCAGCGCGCGAGAGCGCTTCCACTCT 1766  
 QY 152 LysSerGlyTyrLeuValGlyTyrSerCysGlnSerArgIleArgSerTyrLeuArgGluVal 171  
 DB 1765 AAGGAGAGACGTATGAGCCACACAGTCCGACACACGTGTGCGAGGCTTACTAC----- 1715  
 QY 172 SerSerTyrProSerThrValGlyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 188  
 DB 1714 -----TATTAAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1679  
 QY 189 -----ValLeuGlySerMetCysGlnArgGlnArgSer 199  
 DB 1678 TACCGCCAAATAGCCAAAGCCGACAGTGTAAACCTGTGCGAGAGAGAGAGAGAGAGAG 1619  
 QY 200 Met-----GlnTyrAsnGlySerTyrPheAspArgGlyAlaLys----- 212  
 DB 1618 CTGCTCATCGGTGCGACTTTTCTCCATGATGTTTACCGGAAATTGCAAGCAAAAGCAT 1559  
 QY 212 ----- 212  
 DB 1558 GAATTCCTGAAGCAGCACTTGGCGAGCAGAGAAACAGACGCTGGCAGAGATACCAGCAAG 1499  
 QY 213 -----GlyGly 214  
 DB 1498 AGACTGAGCAGAGTGTATGAGGAGTACACCAAGAAATCTGCATCTCGAGAAATGGTCC 1439  
 QY 215 SerArgLeuCysThrProGlnGlyTyrPheSerCysGlnGlyProPhe-----AspMet 232  
 DB 1438 ACTTCCTGTGTCTCCGACTTGGGACTTCTATTTCCAGAGGCTCTTATTCGGAATATGGC 1379  
 QY 233 AspSerCysLeuSerArgHisSerIleAsnProTyrSerAsnArgGlnSerArgIleLeu 252  
 DB 1378 AACAGCTGCTCTAAGCAGACACATATATCCGTAACGCTTCCGCTGAGAAATCTCATTTTG 1319  
 QY 253 PheSerThrTrpAsnLeuAspHisIleIleGlnLysArgThrIleIleProThrLeu 272  
 DB 1318 TTTTCAGGTCTGGAATCTGGACACCAAAATCGAACTGTGCCGACAAATCTTCCGCACTT 1259  
 QY 273 ValGlnAlaIleLysGlu-----GlnAspGly 281  
 DB 1258 GTTGCAAATATGGAGAAATCTTGACATCCGACAGCAAGTGTTCATTTCAATTAAGAG 1199  
 QY 282 ArgGluValAsp-----TrpGlnTyrPheTyrGlyLeuLeuPheThrSerGluAsn 298  
 DB 1198 CAAGTGTGATATCTCAGTACTGAGTACTACTTCTAGAA--AAATCTCCCTTAAGAAAT 1142  
 QY 299 LeuIleValIleHisIleValCysHisLysLys-----ThrThrHis 312  
 DB 1141 CTCAAACTAGTGCACATTTGTCTGTGATGAGAAAGCGGAGCGGCTCAAACTTCGAATGGT 1082  
 QY 313 LysLeuAsnCysAspProSerArgIleTyrLys 323  
 DB 1081 CGCCTACTGTCTCCGACTGTCTCATGAGTATTCGC 1049

XX 14-Aug-2002 (first entry)  
 DT Human CDNA differentially expressed in granulocytic cells #1334.  
 XX  
 DE Human CDNA differentially expressed in granulocytic cells #1334.  
 DE  
 XX Human; ss; granulocytic cell; DNA chip; bacterial infection;  
 KW viral infection; parasitic infection; protozoal infection;  
 KW fungal infection; sterile inflammatory disease; psoriasis;  
 KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
 KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
 KW adult respiratory distress syndrome; inflammatory bowel disease;  
 KW granulocyte activation; chronic inflammation; allergy.  
 KW  
 OS Homo sapiens.  
 PN WO/2002/28999-A2.  
 PD 11-Apr-2002.  
 PD 03-Oct-2001; 2001WO-US30821.  
 PR 03-Oct-2000; 2000US-237189P.  
 PA (GENE-) GENE LOGIC INC.  
 PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
 WP; 2002-435328/46.  
 PT Detecting granulocyte activation by detecting differential expression  
 PT of genes associated with granulocyte activation, which serves as  
 PT diagnostic markers that is useful for monitoring disease states and  
 PT drug toxicity  
 PS  
 XX Claim 1; SEQ ID NO 1334; 114pp; English.  
 XX  
 CC The invention relates to detecting (M1) granulocyte (GC) activation  
 CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
 CC DNA chip analysis as given in the specification, and comparing  
 CC the expression level to an expression level in an unactivated  
 CC GC, where differential expression of Gs is indicative of GCA.  
 CC Also included are modulating (M2) Gs by contacting Gs with an agent  
 CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
 CC for an agent capable of modulating GCA or an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease using the  
 CC gene expression profile; (3) detecting (M4) an inflammation (especially  
 CC chronic) in a tissue, an allergic response in a subject, exposure of a  
 CC subject to a pathogen or sterile inflammatory disease, by detecting the  
 CC level of expression in a sample of the tissue of gene(s) from Gs, where  
 CC the level of expression of the gene is indicative of inflammation;  
 CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
 CC an allergic response in a subject, exposure of a subject to a pathogen  
 CC or sterile inflammatory disease, by contacting a tissue having  
 CC inflammation with an agent that modulates the expression of gene(s)  
 CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
 CC modulating Gs; M3 is useful for screening an agent capable of modulating  
 CC GCA preferably in an inflammation in a tissue; M4 is useful for  
 CC detecting an inflammation (especially chronic) in a tissue, an allergic  
 CC response in a subject, exposure of a subject to a pathogen or sterile  
 CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis,  
 CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
 CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
 CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
 CC periodontal disease, also bacterial infection, viral infection,  
 CC parasitic infection, protozoal infection, fungal infection and M5 is  
 CC useful for treating one of the above conditions. The present  
 CC sequence represents a gene differentially expressed in granulocytes.  
 CC Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1851 BP; 489 A; 525 C; 472 G; 365 T; 0 other;  
 Alignment Scores:  
 Pred. No.: 0.00755 Length: 1851  
 Score: 124.50 Matches: 84  
 Percent Similarity: 34.70% Conservative: 51  
 Best Local Similarity: 21.59% Mismatches: 137  
 Query Match: 6.96% Indels: 118  
 DB: 24 Gaps: 17  
 US-09-748-451-2 (1-338) x ABR84763 (1-1851)  
 OY 4 LysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGly 23  
 625 AAGCCCGGCTG---AGGTGAGGCGGACAACTGGCCGAGACATCATGGCCCTCGGG 681  
 24 ArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuProGluArgLys 43  
 682 ACAAATTCAGG----- 693  
 OY 44 SerArgLeuCysLeuTyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerVal 63  
 694 ---AGAGATGCTTCAGAGAGAG- GAAGCCGAAACACCCCTCAATCTTTCAGACAGGAT 749  
 OY 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTyrGlnGlyTyrValSer 83  
 750 GTTGACAAATGCGCTGCGACGCTTGACCTTGAAAGCGAAAGTGAACTTTGCAAGAA 809  
 OY 84 AspLeuArgTyrPheLeuSerAlaPheHisGluProGlnValAlaGlyLeuGlnAlaAla 103  
 810 GAGATTT---GCCTTTTTCGAAAGAACTCCAGAGAGAAATCCAGAGCTCAGGCTCAG 866  
 OY 104 -----GlnGlnLeuLeuCysAspGluGlnAlaProGlnArgGlnArgLeu 118  
 867 ATTCAGGAAACATGTCATATGATGTGATGTGATGTTCACACCTGACCTCAGCGCTGCC 926  
 OY 119 LeuAlaAspLeuLeuHisAspValSerGlnAsnIleAlaAlaGlnThrArgAlaGluAsp 138  
 927 CTGCGTACGATGCTGACGACATAT---GAAGTGTGCTGCGCAAAACCTGACAGAGCA 983  
 OY 139 ProProTyrPhe----- 143  
 984 GAAGATGTGTACAAATCCAAATGTCGACCTGCTGACGCTGCCACCGAGACAAATGAC 1043  
 OY 144 GlyLeuGluSerArgPheGlnSerLysSerGlyTyrLeuArgTyr----- 158  
 1044 GCCCTGCCCGCAAGACGAGACTCCACTGAGTACCGAGACAGGTGCAGTCCCTCACC 1103  
 159 -----SerCysGluSerArgIleArg----- 165  
 1104 TGTGAAGTGATGCCCTTAAAGAACCAATGAGTCCCTGGAAGCGCAGATGCGTGAATG 1163  
 OY 166 -----SerTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGluAlaGln 183  
 1164 GAAGAGAACTTTCGCGTGAAGCTGTACTACCAAGACACACTATTGGC---CGCCTGAG 1220  
 OY 184 GluGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrAsn 203  
 1221 GATGAGATTCAGAAATATGAGAGGAATGCTGCTGACCTTCGTGAATACCA----- 1274  
 OY 204 GlySerTyrPheAspArgGlyAlaLysGlyLysSerArgLeuCysThrProGluGlyTyr 223  
 1275 -----GACCTG 1280  
 OY 224 PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsnPro 243  
 1281 CTCATGTTAAGATGCGCTTCACATGTGATGATGCGCCACTACAGAACTGCTG----- 1334  
 OY 244 TyrSerAsnArgGluSerArgIleLeu-----PheSerThr----- 255  
 1335 ---GAAGCGAGAGAGAGGATTTCTGCTCTTCCTTCAAACTTTTCTCCCTGAAACCTG 1391

OY 256 -----TyrAsnLeuAspHisIle-----IleGluLysLysArgThrIleIle 269  
 DB 1392 AGGAAACATATCTGATTCACCTCCCTGCTGATACCCACTACAAAGGACATTC--- 1448  
 OY 270 ProThrLeuValGluAlaIleLysGlnGlnAspGlyArgGluValAspTyrPhe 289  
 DB 1449 -----CTGATTAAAGCGTTCGAAACCTGAGATGACAGATTCAC----- 1490  
 OY 290 TyrGlyLeuLeuPheThrSerGluLeuLysLeuValHisIleValCysHisLysLys 309  
 DB 1491 -----GAAACTTCACAGAT-----CAGCATGAC 1514  
 OY 310 ThrThrHisLysLeuAsnCysAspProSerArgIleTyrLysProGlnThrArgLeuLys 329  
 DB 1515 CTGATTAATAAATTCACACACTCAGTGCAGCGGATATATTACCGCAAGATTAATA 1574  
 OY 330 ArgLysGlnProValArgLysArgGln 338  
 DB 1575 AAGAAATCCCATTTCTTAAGAAACAG 1601  
 RESULT 12  
 AAF21607/c  
 ID AAF21607 standard; DNA; 349980 BP.  
 AC AAF21607;  
 DE 13-MAR-2001 (first entry)  
 DE Neisseria meningitidis B nucleotide sequence SEQ ID NO:108.  
 KW Neisseria meningitidis; Neisseria gonorrhoeae; immunogenic; vaccine;  
 KW diagnosis; antigen; detection; infection; gene therapy; antibacterial;  
 KW ds.  
 OS Neisseria meningitidis.  
 PN WO200066791-A1.  
 PD 09-NOV-2000.  
 PE 08-MAR-2000; 2000MC-US05928.  
 PR 30-APR-1999; 99MS-0132068.  
 PR 08-OCT-1999; 99MO-US23573.  
 PR 28-FEB-2000; 2000GB-0004695.  
 PA (CHIR ) CHIRON CORP.  
 PA (GENO-) INST GENOMIC RES.  
 PI Piza M, Hickey E, Peterson J, Tettelin H, Venter JC, Masignani V,  
 PI Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V, Rappuoli R;  
 PI Frazer CM, Grandi G;  
 PI WPI: 2000-647603/62.  
 PS Neisseria meningitidis B full length genome sequence and open reading  
 PS frames are used to detect, treat and prevent Neisserial infections -  
 PS Claim 7; Appendix A; 692pp; English.  
 CC The present invention describes the full length genome of  
 CC Neisseria meningitidis B (NMB). The sequences in AAF21604 and AAF21607  
 CC to AAF21613 represent fragments of the NMB genomic sequence, as the  
 CC sequence was too long to go in a record on its own. It was split into 8  
 CC sequences which overlap each other at the beginning and end of each  
 CC sequence by 49980 bp. The last 49980 bp of AAF21544 is repeated at  
 CC the beginning of AAF21607, and so on). AAF21545 to AAF21607 are repeated the  
 CC Neisseria proteins given in AAF215850 to AAF215893, and AAF21589 to  
 CC AAF21606 represent PCR primers which are used in the amplification of  
 CC the present invention. The NMB genome and fragments from it have  
 CC antibacterial activity, and can be used in vaccines and gene therapy.  
 CC Neisseria nucleic acids, proteins and/or antibodies which binds to the



CC particularly organism-specific probes. Attempts to make efficacious  
 CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
 CC Multivalent vaccines have also been tried but none have successfully  
 CC overcome antigenic variability. The provision of further, complete  
 CC sequences may provide an opportunity to identify secreted or surface  
 CC exposed proteins that may be presumed targets for the immune system and  
 CC which are not antigenically variable or at least more conserved than  
 CC other more variable regions.

XX Sequence 1437668 BP; 344338 A; 353206 C; 385074 G; 355045 T; 5 other;

# Alignment Scores:

Score:	8.37e+03	Length:	1437668
Percent Similarity:	106.50	Matches:	76
Local Similarity:	35.598	Conservative:	45
Match:	22.358	Mismatches:	128
	5.958	Indels:	97
	21	Gaps:	17

US-09-748-451-2 (1-338) x AAA81490 (1-1437668)

```

OY 14 ArgSerProArgLysPheGly---ValAlaGlyArgSerCysGlnGluValLeuArgLys 32
DB 386372 CAACCCCGCAAAACGGGGGTTGTCACATCATCATGACCGACGCTTTCACGAA 386313
OY 33 GlyCys-----LeuArgPheGlnLeuProGluArgGlySerArgLeuCysLeu 48
DB 386312 TATTGCACAGTACCTGCTGACAAATCCACTAACCTCATGATCGGCAAGATATGCTTA 386253
OY 49 TyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerValProAspAsnAlaGlu 68
DB 386252 -----GCCTTAGTG 386244
OY 69 LeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGlyTyrValSerAsp-IleArgArgPh 88
DB 386243 TTGCTGATGATGTTGCGCGGACACACCTGACGGGTTAC-----GATGCTGCGGATTTT 386190
OY 88 eleuSerAlaPheHisGluProGlnArgGlnAlaGlyLeuIleGlnAlaGlnGlnLeuCys 108
DB 386189 CACATCCACACGGCGGTGAGTTCGATATCATCATGACATGAGGCTCGCTTTTGGCTTT 386130
OY 108 sAspGlnGlnAlaProGlnArgGlnArgGlnAlaAsp-LeuLeuHisAsnValSerG 128
DB 386129 AGAGACTTTCGACCCAGT-----TTGGCACTTCGGCTTCACAAACTT---G 386085
OY 128 LnsnIleAlaAlaGluThrArgAlaGluAspProProtrPheGlnGlyLeuGlnSerA 148
DB 386084 AGTCATTTGGCTTGAGATTGCG-----CCAGAGCC 386055
OY 148 rGpHeGlnSerLysSerGlyTyrLeuArgTyrSerCysGlnSerArgIleArgSerTyrL 168
DB 386054 GACAGCAGAAATTCAGAAACAGATACCCCTGCTGACACGAGTTGTCACACTAGCTA 385995
OY 168 euArgGluValSerSerTyrProSerThrValGlyAlaGluAlaGlnGlnGluPheLeuA 188
DB 385994 TCGCGCGGCTTCGAGAGGCGCTGATGTAT----- 385963
OY 188 rGValLeuGlySerMetCysGlnArg-----LeuArgSerMetGlnT 202
DB 385962 -----TGTCAAGACCCATTAAGTCGATGTCGCTTCGACCGGACAG 385920
OY 202 yrasnGlySerTyrPheAspArgGlyAlaGlyAlaGlySerArgLeuCysThrProGlnG 222
DB 385919 TT---GGACATTTGTTGCGCTGACGCGT----- 385894
OY 222 LyrTrpPheSerCysGlnGlyProPheAspMetLaspSerCysLeuSer---ArgHisSerI 241
DB 385893 -----GGCGCAATR---CTTTCAGGTGTGCTTTCGAGCTTCGCGCGCA 385854
OY 241 leasnProTyrSerAsnArgGlnSerArgIleLeuPheSerThrTrpAsnLeuAspHisI 261
DB 385853 TGAATCTTATGCAACCGAAGAGGCTTTGGCAGACAGGAAATGCTTTCATCAACATR 385794

```

```

OY 261 lelleGlu-----LysIleArgThrIleleProThrLeuValGluA 275
DB 385793 TTGCGAGCGCTTCCTCACAAACGACAGAGCGGGGTGCGGCTTCGCTGCGG 385734
OY 275 lalleGlnGlnAspGlyArgGluValAspTrpGlnTyrPheTyrGlyLeuLeuPheT 295
DB 385733 TTGCGGGGACACACCGCATCGCGGATTCCTACGAGAC-----CTTGCGCTGCTTATC 385680
OY 295 hrSerGluAsnLeuLysLeuVal---HisIleValCysHisLys---LysThrThrHisL 313
DB 385679 AAAGTACGGCGTTTTCACAGATTCGCCATTCGCTGCTGATTCGATTCGACATGCGCC 385620
OY 313 yslenuAsnCysAspProSerArgIleTyrLysProGlnThrArgLeuLysArgLys 331
DB 385619 GCTTACGGTATACCTGACAGGACGACGCGCTCGCCGACAGACATGCGGAAAGCA 385564

```

RESULT 14  
 ID AAA81540 standard; DNA; 782 BP.  
 AC AAA81540;  
 XX 04-DEC-2000 (first entry)  
 XX N. meningitidis partial DNA sequence gnm\_87 SEQ ID NO:87.  
 XX Neisseria meningitidis; Neisseria gonorrhoeae; genome; immunogenic;  
 XX antigen; vaccine; diagnosis; infection; antibacterial; identification;  
 XX Meningococcus B; MenB; ds.  
 XX Neisseria meningitidis.  
 XX WO200022430-A2.  
 XX 20-APR-2000.  
 XX 08-OCT-1999; 99WO-US23573.  
 XX 09-OCT-1998; 98US-0103794.  
 XX 30-APR-1999; 99US-0132068.  
 XX (CHIR ) CHIRON CORP.  
 XX Frizer CM, Hickey E, Peterson J, Tettein H, Venter JC;  
 XX Maignant V, Galeotti C, Mora M, Ratti G, Scarselli M, Scarlato V;  
 XX Rappuoli R, Pizza M;  
 XX WPT: 2000-318079/27.  
 XX Isolated nucleotide sequences of Neisseria meningitidis which can be  
 XX used in the diagnosis and treatment of N. meningitidis infection and  
 XX other Neisserial infections, for example, N.gonorrhoea -  
 XX Claim 7: Page 1482-1483; 1760pp; English.  
 XX The present invention describes methods of obtaining immunogenic  
 XX proteins from Neisseria genomic sequences. AAA81453 to AAA82414  
 XX represent specifically claimed Neisseria meningitidis genomic DNA  
 XX sequences; AAA81260 to AAA81303 and AAA825620 to AAA825663 represent  
 XX Neisseria DNA sequences and their corresponding proteins; AAA81254 to  
 XX AAA81259 and AAA81304 to AAA81321 represent PCR primers used in the  
 XX isolation of Neisseria meningitidis DNA sequences; and AAA8122 to  
 XX AAA81452 represent Neisseria meningitidis MenB polynucleotide ORF  
 XX sequences, which are all used in the exemplification of the present  
 XX invention. The nucleic acid sequences, protein sequences, and antibodies  
 XX against them, can be used in the manufacture of a composition. The  
 XX composition can be used as a medicament (or in the manufacture of a  
 XX medicament) for treating, preventing or diagnosing infection due to  
 XX Neisserial bacteria. For example, some of the identified proteins could  
 XX be components of vaccines against Meningococcus B; against all serotypes;  
 XX and/or against all pathogenic Neisseriae. Identification of sequences  
 XX from the bacterium will also facilitate production of biological probes,  
 XX particularly organism-specific probes. Attempts to make efficacious

CC Meningococcus B vaccines have failed mainly due to antigen tolerance.  
CC Multivalent vaccines have also been tried but none have successfully  
CC overcome antigenic variability. The provision of further, complete  
CC sequences may provide an opportunity to identify secreted or surface  
CC exposed proteins that may be presumed targets for the immune system and  
CC which are not antigenically variable or at least more conserved than  
CC other more variable regions.

XX SQ Sequence 782 BP; 200 A; 205 C; 198 G; 179 T; 0 other;

#### Alignment Scores:

Pred. No.:	0.238	Length:	782
Score:	105.00	Matches:	46
Percent Similarity:	37.88%	Conservative:	29
Best Local Similarity:	23.23%	Mismatches:	69
Query Match:	5.87%	Indels:	54
DB:	21	Gaps:	11

US-09-748-451-2 (1-338) x AAA01540 (1-782)

168 LeuArgGluValSerSerTyrProSerThrValGluAlaGluGluPhe-- 186

658 CTTCAACAACTGATGATTTGCGCTGAGATTCGCCCAAGCCGACAGCAAAATTCAG 599

187 -----LeuArgValLeuGly----- 191

598 AAACAGGTACGCGCTTCTGCACAGGTTGTTGCCACTAGTCCGCGGCTCTGACA 539

192 -----SerMetCysGlnArg-----LeuArgSerMetGlnTyr 202

538 GGGCGCTGATGATTTGTTGCAACCAATTAAGTGGATGCGCTTCACAGCGGACAGATT 479

203 AsnGlySerTyrPheAspArgGlyAlaLysGlySerArgLeuCysThrProGluGly 222

478 ---GGACATTGCTGCTGACCTGAGCGGCTT----- 455

223 TrpPheSerCysGlnGlyProPheAspMetAspSerCysLeuSer---ArgHisSerIle 241

454 -----GGCCCAATA---CTTCAAGATTGTCTTGAGCTTCGCGCGCAATG 413

242 AsnProTyrSerAsnArgGluSerArgIleLeuPheSerThrTyrAsnLeuAspHisIle 261

412 AATCCCTATCGAAGAACCGAACAGGCTTTGGCAGACAGGAAATGTTTCATCAACATATT 353

262 IleGlu-----LysLysArgThrIleIleProThrLeuValGluAla 275

352 GCGAGAGCTGCTCCACACAGACAGACAGCGCGGCTCGGCTTGGCTCGGCTT 293

276 IleLysGluGlnAspArgGluValAspThrIleTyrPheTyrGlyLeuLeuPheThr 295

292 CGGGGCGACAGACCGCATCGCGCATCTACGCGAC-----CTTGGCTGCTTATCA 239

296 SerGluAsnLeuLysLeuVal---HisIleValCysHisLys---LysThrThrHisLys 313

238 AGTAGCGGCTTTTCCAGCATTCGCCATAGTGTGGCTAGCATATTCGTCAGTCCGCGC 179

314 LeuAsnGlyAspProSerArgIleTyrLysProGlnThrArgLeuLysArgLys 331

178 TTACGGGTGAACCTGTAACGAGACGCGTCCGCCACACAGACAGTCCGGAACGGA 125

#### RESULT 15

AAA09307 ID AAA09307 standard; DNA; 1766 BP.

AAA09307; 10-AUG-2000 (first entry)

XX Human vimentin coding sequence.

DE NIK1 interacting Protein; vimentin; protein complex; cytoskeletal;

KW antiviral; neuroprotective; cardiant; ss.

XX

OS Homo sapiens.

XX Key Location/Qualifiers

XX CDS 44..1444

XX FT /\*tag= a

XX W0200020448..42.

XX 13-APR-2000.

XX 06-OCT-1999; 99WO-US23314.

XX 06-OCT-1998; 98US-0167206.

XX (CURA-) CURAGEN CORP.

XX Nandabalan K, Schulz VP, Yang M;

XX WPI, 2000-303742/26.

XX P-PSDB; AAY92335.

XX New complex of a NIK1 protein and a NIK1 protein-interacting protein,

XX useful for treating cancer, hyperproliferative disorder,

XX neurodegenerative disorder, cardiomyopathies, viral infections and

XX metabolic disorders

XX Example 1; Page 142-145; 172pp; English.

XX AAA09303-09 were isolated in a modified yeast two hybrid system using

XX NIK1 protein as "bait". These are known sequences which encode NIK1

XX interacting proteins. The invention concerns purified complexes of a

XX NIK1 protein and a NIK1 protein-interacting protein, where the

XX interacting protein is chosen from TrkA, protein phosphatase 1alpha,

XX 14-3-3epsilon, alpha-tropomyosin, vimentin, p0071, Irf-1, Ip-1, Ip-2,

XX Ip-3, Ip-4, or Ip-5. NIK1 (also referred to as Nek2) is a human

XX homologue of the Aspergillus nidulans mitotic regulator, NIMA kinase.

XX NIK1 is a serine/threonine-specific kinase and is thought to play a key

XX role in cell-cycle events leading to the onset of mitosis. The

XX complexes, their derivatives and NIK1 or NIK1-IP protein and DNA

XX sequences, etc. are useful for treating or preventing a disease or

XX disorder involving aberrant levels of the complex or protein. Such

XX disorders include cancer, hyperproliferative disorders, viral

XX neurodegenerative disorders, cardiomyopathies, viral infections and

XX metabolic disorders.

XX Sequence 1766 BP; 479 A; 482 C; 449 G; 356 T; 0 other;

XX SQ

XX Alignment Scores:

XX Pred. No.: 1.4 Length: 1766

XX Score: 102.50 Matches: 77

XX Percent Similarity: 34.42% Conservative: 50

XX Best Local Similarity: 20.87% Mismatches: 128

XX Query Match: 5.73% Indels: 115

XX DB: 21 Gaps: 16

XX US-09-748-451-2 (1-338) x AAA09307 (1-1766)

XX 4 LysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGly 23

XX 546 AAGCCGCGCTGCG---AGGTGGAGCGCGCAACACCTGCGCGGAGACATCATCGCGCGG 602

XX 24 ArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuProGluArgGly 43

XX 603 AGAATTGACAGG----- 614

XX 44 SerArgLeuSerValTyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerVal 63

XX 615 ---AGGAGATGCTTCAACAGAGAG-GAAGCCGAAACACCTCGCATCTTTCACAGCAGAT 670

XX 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaIleArgGlnGlyTyrValSer 83

XX 671 GTTGACAAATGCGCTCTGCGACCGCTTGACCTTGACCAAGAGCAATCTTGCAAGAA 730



OY		119	LeuAlaSerLeuValAsnValSerGlnAsnIleValIleAlaLysThrArgAlaGluLysP	138
Dd		1009	CTGCSTGACAGTACCGCAGCAATAT---GAAGGTGGCTGCCAAGAACCTGCAGGAGCCA	1065
OY		139	ProPrThrPhe-----Glu	143
Dd		1066	GAAAGAATGGTACAAATCCAGATTGCTGACTCTCTGAGGCTGCCAACCGAACATATAC	1122
OY		144	GlyLeuGluSerArgPheGlnSerIlysSerGlyTyrrLeuAlaGtyr-----	158
Dd		1126	GCCCTGCCCCAGCAACCGAGAGTCCTCAGTAGTACCGGACAAGTCAGTCCCTCACCC	1185
OY		159	-----SerCysGluSerArgIleArg-----	165
Dd		1186	TGTGAAGTGATGCCCTTAAGAGAACCAATAGTCCCTCGGAAGCCCAATGCGTGAATATG	1245
OY		166	-----SerTyrrLeuArgGluValSerSerTyrrProSerThrValGlyAlaGluAlaGln	183
Dd		1246	GAAGAGAACTTTGCCGCTGAAGCTGCTCACTACCAAGACATAATATGCG---CGCCTGACG	1302
OY		184	GluGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrrAsn	203
Dd		1303	GATGAGATTCAGAAATATGTAAGAGAGAAATGGCTGCTGCATCCCTCGTAATACCAG---	1356
OY		204	GlySerTyrrPheArgPheArgGlyAlaValSerGlySerArgLeuCysThrProGluGlyTrp	223
Dd		1357	-----GACCTG	1367
OY		224	PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleasnPro	243
Dd		1363	CTCAATGTTAGATGGCCCTTGACATTCAGATTGGCACTACAGAACCTGCTG---	1416
OY		244	TyrSerAsnArgGluSerArgIleLeu-----PheSerThr-----	255
Dd		1417	--GAAGCGAGAGAGACGAGATTTCTCTGCTCTCCAAACTTTCTCCCTCCCTGAACTCG	1473
OY		256	-----TrpAsnLeuAspHisIle-----IleGluLysArgThrIleLe	269
Dd		1474	AGGGAACCTAATCTGGATTCACCTCCCTTGCTGATACCCACCTCAAAAAGGACACTT---	1530
OY		270	ProThrLeuValGluAlaIleLysGluGlnAspGlyArgGluValASP-----	285
Dd		1531	-----CTGATTAAAGACGGTTGAACACTGAGANTGACGGTTATCAACGAACCTCTCAG	1584
OY		286	-----TrpGluTyrrPheTyrrGlyLeuLeuPhe--ThrSerGI	297
Dd		1585	CATCAGCATGACCTTGATATAAAAAATGGCACACACTCACTGACAGATATATATTACACGAA	1644
OY		297	uasnLeuLysLeuValHisIleVal	305
Dd		1645	GAATTAATAAAGAAATCCATATCTTA	1669
XN		17		
XN		122		
XX		AAT04122 standard; DNA: 1209 BP.		
AC		AAT04122;		
XX		09-MAY-1996 (first entry)		
DE		Porcine acylglucosamine-2-epimerase DNA.		
XX				
KW		Porchee; acylglucosamine-2-epimerase; N-acetylmannosamine;		
XX		N-acetylneuraminic acid; renin-binding; enzymatic production; ss.		
OS		Sus scrofa.		
XX				
FH		key	Location/Qualifiers	
FT		CDS	1..1209	
XX			/tag= a	
XX		W09526399-A1.		

```

Dc 552 GCTGTCGTGGCTGGTGTAGACAGCTGGGGGAGAGACGAGACGATGTCGAGCCGCTACGC 61111
Oy 205 eRTYrPhAsPAsrGgLYaLYaLYaSGLYgYSerAryLeuCYstrProLugLYrPhPhe 224
Db 612 GCAGCTGGGGGACATGGTGTGGCTCGGAGGATCC---TTCACACAGCTCCAGAGGATGAGCA 666
Oy 225 SerCYgLnGlyPro-Phe-----AspMetLAsPserCYtLeuSerArGI 239
Db 669 GGCTGTCTGTGGAGAAATGTGTCCGGAAGATGGCGAGAGAACTTTCGGCTGCTGGGAGACA 728
Oy 239 sSerLLeAsnPro-----TyrSerAsnAr 247
Db 729 CCAG---AACCCAGGACGACGCGCTGGAACTGTGCTTCTCTCTCCGACAGCAGCG 785
247 g-----GluSerArgLLeuLeuPhe----- 253
786 GAGCGGTGACGCCAAACTTCGAGCCACGTCATGCACAGCTTCTGTACTGCTTCCTCCG 845
Oy 254 -SerThrTrpAsnLeuAspHis 260
Db 846 CTCGCGATGGAGCGCTGATCAC 867
RESULT 18
ABN51606
ID ID ABN51606 standard; DNA; 65 BP.
XX
XX ABN51606;
AC
AD
DE 15-JUL-2002 (first entry)
XX
XX Mouse spliced transcript detection oligonucleotide SEQ ID NO:24354.
XX
XX Human; mouse; ref; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.
XX
XX Mus musculus.
XX
XX NO200210449-A2.
XX
XX 07-FEB-2002.
XX
XX 20-JUL-2001; 2001WO-IB01903.
XX
XX 28-JUL-2000; 2000US-221607P.
XX
XX 02-MAY-2001; 2001US-287724P.
XX
XX (COMP-) COMPUGEN INC.
XX
XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX
XX WPI; 2002-257383/30.
XX
XX
XX New oligonucleotide libraries comprising oligonucleotides which
XX selectively hybridize to mRNAs transcribed from a transcription unit of
XX a genome, useful for detecting tissue-, pathology-, and
XX developmental-specific genes -
XX
XX Example 1; SEQ ID 24354; 47pp; English.
XX
XX
XX The present invention describes oligonucleotide libraries for detecting
XX messenger RNAs that populate a (sub-)transcriptome, where the
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX transcription units that populate a genome. The library comprises
XX several oligonucleotides, each capable of hybridizing selectively to a
XX set of messenger RNAs transcribed from a given transcriptome unit of
XX the genome, which encodes one or more messenger RNA splice variants.
XX The oligonucleotide libraries are useful for detecting mRNAs from a
XX biological sample, in expression profiling studies, in qualitatively or
XX quantitatively characterizing the corresponding transcriptome, and in
XX detecting RNA transcripts and splice variants of human or animal
XX transcriptomes. The libraries may also be used as specialised mint
XX libraries to detect transcripts of a sub-transcriptome under a

```

CC particular biological or pathological state, and so allowing the  
CC detection of tissue- and pathology-specific genes such as those genes  
CC only expressed in specific tissue under a specific pathological  
CC condition; to detect developmental specific genes; and to detect RNA  
CC transcripts and splice variants of a transcriptome of a patient suffering  
CC from a particular disorder. ABN72253 to ABN5589 represent  
CC oligonucleotide sequences from rats, humans and mice, which are used in  
CC the exemplification of the present invention.  
CC N.B. The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

XX  
XX Sequence 65 BP; 15 A; 18 C; 17 G; 15 T; 0 other;

Alignment Scores:  
Pred. No.: 0.0173 Length: 65  
Score: 101.00 Matches: 18  
Percent Similarity: 95.24% Conservative: 2  
Best Local Similarity: 85.71% Mismatches: 1  
Query Match: 5.65% Indels: 0  
DB: Gaps: 24

US-09-748-451-2 (1-338) x ABN51606 (1-65)

OY 284 ValAspTrpGluTyrPheTyrGlyLeuPheRserGluAsnLeuTysLeuValHis 303  
Db 3 GTGACCTGGAGTACTCTCTACAGCCGCTCTTCACATCGCAGAACTGAAGCTGGTGAC 62

OY 304 Ile 304  
Db 63 ATC 65

RESULT 19  
ABI99777  
ID ABI99777 standard; cDNA; 2174 BP.

AC ABI99777;  
XX  
XX 07-MAR-2002 (first entry)  
DX  
XX Mouse ischemic condition related cDNA sequence SBO ID NO:866.  
XX  
XX Mouse: Ischaemia; compressive Ischaemia; occlusive Ischaemia;  
XX  
XX vasospastic Ischaemia; Ischaemic condition; Ischaemic disease; ss.  
XX  
XX Mus musculus.  
OS  
XX WO200188188-A2.  
PN  
XX 22-NOV-2001.  
PD  
XX 18-MAY-2001; 2001WO-JP04192.  
XX  
XX 18-MAY-2000; 2000JP-0145977.  
PR  
XX (UYNI-) UNIV NIHON SCHOOL JURIDICAL PERSON.  
PA  
XX  
XX Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;  
PI  
XX WPI; 2002-034733/04.  
DR  
XX P-PSDB; ABB57310.  
XX

PT Examining the ischemic condition (e.g. occlusive Ischemia) by measuring  
PT expression levels of particular genes defined in the specification, or  
PT by determining the expression profile of a gene group comprising these  
PT genes -  
XX  
XX Claim 2; Page 2174-2177; 2690pp; English.  
XX  
XX The present invention describes a method for examining ischemic  
XX conditions, comprising measuring the expression levels of particular  
XX genes (1) in a test sample or determining the expression profile of a  
XX gene group in the sample comprising genes selected from (1). The method



CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
XX Sequence 7545 BP; 2136 A; 1578 C; 1563 G; 2268 T; 0 other;  
SQ

Alignment Scores:

Pred. No.:	33.6	Length:	7545
Score:	98.00	Matches:	72
Percent Similarity:	34.04%	Conservative:	56
Best Local Similarity:	19.15%	Mismatches:	132
Query Match:	5.48%	Indels:	117
DB:	23	Gaps:	16

US-09-748-451-2 (1-336) x ABUL4002 (1-7545)

QY 18 LysPheGlyValAlaGlyArgSerCysGlnGluValLeuArgGlyGlyCysLeuArgPhe 37  
|||  
3174 MAGAGCGTCTCTCGCGCTCCAGTCTGGAGAGATCCGAGCAAGTGGCGGAGAACTTC 3115  
|||  
38 GlnLeuProGlnArgGlySerArgLeuCysLeuTyGlnAspGlyThrGlnLeu---Thr 56  
|||  
3114 GAGAAAGTGGACACCTGGCCACCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGT 3055  
|||  
QY 57 GluAspTyrPheProSerValProAspAsnAlaGluLeuValLeuThrLeuGlyGln 76  
|||  
3054 GAGAGTACTTCTCGACCTCGAGCAAGATAGCGAATGTGGCCGCTTTCCGGAGAA 2995  
|||  
QY 77 AlaTrp-----Gln 79  
|||  
2994 CATGTGATCGATGTGAGTACCCCTTAATTAAGAAAGAAAGAAACAGCACTTAATAAT 2935  
|||  
QY 80 GlyTyrValSer-----AspIleArg 86  
|||  
2934 AATTATATACGACCTTTTATTATTAATCACCCTCAATTAATTCGTTTGGCATTTGTGTA 2875  
|||  
QY 87 ArgPheLeuSerAlaPheHisGluProGlnValGly---LeuIleGlnAlaGlnGln 105  
|||  
2874 ATTTATATTCTCAAGAGAGACACCGGTATATGATGATCTATTATGTGCTCTCT 2815  
|||  
QY 106 LeuLeuCys-----AspGlu 110  
|||  
2814 TTTATCTGCATTAAGCATATATTCATTTCTTGATGCCATGAAGAGCTCCATGAG 2755  
|||  
QY 111 GlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuLeuHisAsnValSerGlnAsnIle 130  
|||  
2754 CATGGACCAAGCCAGAGAGATT-----AAATTAAATGCTTATCTGCACCTCTGTT 2704  
|||  
131 AlaAlaGluThrArgAlaGluAsp-----ProProTyrPheGluGly-----LeuGlu 146  
|||  
2703 ATAGCCCACTCTAGTACGATACCACTCCACATGGCAAGGAGGAGGAGGAGGAGGAGGAG 2644  
|||  
QY 147 SerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerArgIleArgSer 166  
|||  
2643 CGGAGAGCTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2596  
|||  
QY 167 TyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGluAlaGlnGluPhe 186  
|||  
2595 -----CTCGAGATGGCTCGCATCCGAGTGGTGGGCGGAGCTGCAGAACACACCTCTG 2542  
|||  
QY 187 LeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrAsn----- 203  
|||  
2541 CAAGGTGTCGATGAGAGATGCCGATCGAGTGGACTCTCTCCACATGATGCCACTC 2482  
|||  
QY 204 ---GlySerTyrPheAspArgGlyAlaLysGlySerArg----- 216  
|||  
2481 GTTGTGGACATCAGCGGAAAGAGTTTATGAGACAGCTCAAGAGATCGGGAAGGTGAGT 2422  
|||  
QY 217 -----LeuCys 218  
|||  
2421 TTAATTCCTTCCCATATACCACTTAATGTAGAGATTTTCATGACTTTAGCCATTGT 2362  
|||  
QY 219 ThrProGlnGlyTyrPheSerCysGlnGlyProPheAspMetLysSerCysLeuSerArg 238  
|||  
2361 GTGCAAGCGGAATG-----CCGAGGATCGTTAAATCTGCTGAAGCTAT 2317

QY 239 HisSerIleAsnProTyrSerAsnArgGluSerArgIleLeuPheSerThrTrpAsnLeu 258  
|||  
DB 2316 TGAAGCTGTGGCATTTCT----- 2296  
|||  
QY 259 AspHisIleIleGluLysArgThrIleIleProThrLeuValGluAlaIleLysGlu 278  
|||  
DB 2295 -----GCTCGAAGCTTATCCGAGAGATGCG-GAGGCCATTGACCG 2255  
|||  
QY 279 GlnAspGlyArgGluValAspTyrGluTyrPheTyrGlyLeuLeu-----PheThrSer 296  
|||  
DB 2254 GAGATTGGACGCGAGTTGAAC---GAGGCGAGAGTGGACAGATGAGCACCACACGACG 2198  
|||  
QY 297 GluAsnLeuLysLeuValHisIleVal---CysHisLysLysThrThr 311  
|||  
DB 2197 AGCAACACCCGACCATCGAGTGTGTCAGTGGCAATCAAAACAG 2150  
|||

Search completed: May 26, 2003, 16:31:19  
Job time : 729 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 26, 2003, 15:24:28 ; Search time 63 Seconds  
(without alignments)  
1645.345 Million cell updates/sec

Title: US-09-748-451-2  
Perfect score: 1/89  
Sequence: 1 MLDKPKSVKRLAIRSPKRC.....SKIKPQTRIKRQPKVKKRQ 338

ling table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 44362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n model -DEV-xlh  
-O-/cgn2\_1/USPTO.spool/US09748451/unat.21052003.153830.16303/app.query.fasta\_1.519  
-DB-issued Patents.NA -OPMT-fastap -SUFFIX-trn1 -MINMATCH=0.1 -LOOPCL=0  
-LISTEX=0 -UNITS-bits -SPAR=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdl  
-LIST=45 -DOCLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=20  
-MODE-LOCAL -OPMT=pcr -MODE=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPT=US09748451.ecgn.1.1.40.ernat.21052003.153830.16303 -NCPU=6 -ICPU=3  
-NO\_MMAP -IAREGUTERY -NEG SCORES=0 -WAIT -DSPROCK=100 -LONGIOG  
-DEV.TIMEOUT=120 -MARN.TIMEOUT=30 -THREADS=1 -GAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:  
Issued Patents.NA:  
1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1789	100.0	2839	4	US-09-061-702-1
2	101.5	5.7	1209	1	US-08-553-703A-5
3	101.5	5.7	1209	1	US-08-553-703A-9
4	101.5	5.7	1209	2	US-09-006-021-5
5	101.5	5.7	1209	2	US-09-006-021-9
6	96	5.4	2729	1	US-07-938-782A-1
7	96	5.4	2729	1	US-08-630-524-1
8	96	5.4	2729	5	PCr-US93-08331-1
9	94	5.3	2143	4	US-09-083-023-21
10	91	5.1	1167	4	US-09-083-023-19
11	89.5	5.0	1301	4	US-09-083-199B-8
12	89	5.0	1689	4	US-09-061-702-3

13	88.5	4.9	9633	6	US-09-069-023-13	Patent No. 5223423
14	88	4.9	1049	4	US-08-794-158-1	Sequence 13, Appl
15	88	4.9	2637	4	US-08-794-158-1	Sequence 1, Appl
16	88	4.9	2637	4	US-08-794-158-3	Sequence 3, Appl
17	86.5	4.8	4322	2	US-08-537-342-1	Sequence 1, Appl
18	86	4.8	1114	4	US-09-069-023-17	Sequence 17, Appl
19	85.5	4.8	4403765	4	US-09-103-840A-2	Sequence 2, Appl
20	85.5	4.8	4411529	4	US-09-103-840A-2	Sequence 1, Appl
21	85	4.8	1155	1	US-08-707-792A-2	Sequence 2, Appl
22	85	4.8	1155	1	US-08-707-792A-2	Sequence 1, Appl
23	84.5	4.7	1124	3	US-09-329-884-13	Sequence 13, Appl
24	84.5	4.7	1124	3	US-08-945-056-1	Sequence 4, Appl
25	84.5	4.7	4201	3	US-08-945-056-1	Sequence 1, Appl
26	84.5	4.7	15101	2	US-08-799-464A-14	Sequence 14, Appl
27	84.5	4.7	15101	5	PCT-US95-09927-14	Sequence 14, Appl
28	84.5	4.7	15108	4	US-08-157-005-1	Sequence 1, Appl
29	84.5	4.7	15108	4	US-08-747-863-1	Sequence 1, Appl
30	84.5	4.7	15108	4	US-09-565-664-1	Sequence 1, Appl
31	84	4.7	4287	1	US-08-244-189-1	Sequence 1, Appl
32	84	4.7	4287	1	US-08-306-691B-53	Sequence 53, Appl
33	84	4.7	20235	1	US-07-642-734C-3	Sequence 3, Appl
34	84	4.7	20235	3	US-08-439-009A-3	Sequence 3, Appl
35	83.5	4.7	2680	2	US-08-533-306A-3	Sequence 5, Appl
36	83.5	4.7	2680	2	US-08-742-923A-5	Sequence 5, Appl
37	83.5	4.7	4403765	4	US-09-103-840A-2	Sequence 2, Appl
38	83.5	4.7	4411529	4	US-09-103-840A-1	Sequence 1, Appl
39	82.5	4.6	2268	2	US-08-873-093-2	Sequence 2, Appl
40	82.5	4.6	2268	2	US-09-206-646-2	Sequence 2, Appl
41	82.5	4.6	3168	4	US-08-659-251-14	Sequence 14, Appl
42	82.5	4.6	3168	4	US-09-256-490-14	Sequence 14, Appl
43	82.5	4.6	3168	5	PCT-US96-11445-14	Sequence 14, Appl
44	82.5	4.6	10163	2	US-08-659-251-1	Sequence 1, Appl
45	82.5	4.6	10163	4	US-09-256-490-1	Sequence 1, Appl

#### ALIGNMENTS

US-09-061-702-1  
Sequence 1, Application US/09061702  
Patent No. 6165737  
GENERAL INFORMATION:  
APPLICANT: Wang, Xiaodong  
APPLICANT: Liu, Xuesong  
TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN  
TITLE OF INVENTION: APOPTOSIS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESSES:  
ADDRESS: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/061,702  
FILING DATE: Concurrently Herewith  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: McMillian, Nabuela R.  
REGISTRATION NUMBER: P-43,363  
REFERENCE/DOCKET NUMBER: US95,546  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512)418-3000  
TELEFAX: (512)418-7577  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2839 base pairs

TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-061-702-1

## Alignment Scores:

Pred. No.:	6,556-208	Length:	2839
Score:	1789.00	Matches:	338
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	4	Gaps:	0

US-09-748-451-2 (1-338) x US-09-061-702-1 (1-2839)

```

QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20
DB 132 ATGCTCCAGAAAGCCAGAGCGTGAACCTGGGCCCTGGGACCCGAGAGAGTTGGC 191
QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40
DB 192 GTGGCTGGCCGAGAGCTCCAGAGGTGCTGCGCAAGGGCTGTCTCCCTCCAGCTCCCT 251
QY 41 GluArgGlySerArgLysCysLeuTyrGluAspGlyThrGluLeuThrGluAspTyrPhe 60
DB 252 GAGCGCGGTTCGCGCTGTGCTGTACGAGAGATGCGAGAGCTGAGAGGAGATTACTTC 311
QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaArgPheGlnGly 80
DB 312 CCCAGTGTCCGCAACAGCCGAGCTGTGCTGCTACCTTGGGCCGAGGCTGGCAGAGGC 371
QY 81 TyrValSerAspLeuArgPheLeuSerLysPheLysGlnProGlnValGlyLeuLeu 100
DB 372 TATGTGAGCAGATCAGAGCGCTTCCAGATTCACAGAGCCAGAGGAGGCGCTATC 431
QY 101 GlnAlaAlaGlnGlnLeuLeuCysAspGlnGlnAlaProGlnArgGlnArgLeuLeuAla 120
DB 432 CAGGCGCCGAGAGCTGTGTGTGTGATGACAGAGCCCAAGAGGAGAGGCTGTGCT 491
QY 121 AspLeuLeuHisAsnValSerGlnAsnIleAlaAlaGlnThrArgAlaGluAspProPro 140
DB 492 GACCTCTGCAACAGCTCAGAGACATCGCGGCCGAGAGCGGCTGAGAGACCCGCC 551
QY 141 TrpPheGlnGlnLeuGlnSerArgPheGlnSerLysSerLysGlyTyrLeuArgTyrSerCys 160
DB 552 TGGTTTGAAGGCTTGGAGTCCCATTTCAAGAGCAAGCTGCGATTCGACATACAGCTGT 611
QY 161 GluSerArgLysLeuArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAla 180
DB 612 GAGAGCGCGATCCGAGATTACCTGAGGAGAGTGAAGCTTACCCCTCCACAGTGGTGG 671
QY 181 GluAlaGlnGlnGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMet 200
DB 672 GAGGCTCAGAGAGATCTGCGGGTCTCGGCTCCATGTGCCAGAGGCTCCGGTCCATG 731
QY 201 GlnTyrAsnLysSerTyrPheAspArgLysAlaLysGlyLysSerArgLeuGlySerPro 220
DB 732 CAGTACATAGGCGAGCTTCTGACAGAGAGCCAAAGGCGGACCCCTCTGCACAGC 791
QY 221 GlnGlyTyrPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240
DB 792 GAAGCGCTGCTTCTGCGAGGCTCTTGACATGAGACAGCTTATATCAAGACATCC 851
QY 241 IleAsnProTyrSerAsnArgLysSerArgLysLeuPheSerThrTrpAsnLeuAspHis 260
DB 852 ATCAACCCCTTACAGTACAGAGAGAGAGATCCTTTCAGACCTGGAACCTGGATCAC 911
QY 261 IleIleGlnLysArgThrIleIleProThrLeuValGlnAlaIleLysGlnGlnAsp 280
DB 912 ATATATGAAAAAGAGCCACCATCATCTTACACTGTGTGAAGCAATTAAGAAACAGAT 971
QY 281 GluArgGluValAspTyrGlnTyrPheTyrGlnLeuLeuPheThrSerGlnAsnLeuLys 300

```

```

DB 972 GGAAGAGAGATGAGCTGGAGTATTTTATGAGCCCTGTTTAACTTCAGAAACCTTAA 1031
QY 301 LeuValHisIleValCysHisLysLysThrThrHisLysLeuAsnCysAspProSerArg 320
DB 1032 CTAGTGCACATTTGCTGCTCCATTAAGAAAAACCAACGCTCAACTGTGACCCGAGAGA 1091
QY 321 IleTyrLysProGlnThrArgLeuLysArgLysGlnProValArgLysArgGln 338
DB 1092 ATCTAACAAACCCAGACAGAGTTGAAGCGGAGAGGCTGTGGGAAACCCAG 1145

RESULT 2
US-08-553-703A-5
Sequence 5, Application US/08553703A
Patent No. 5795767
GENERAL INFORMATION:
APPLICANT: MARU, ISAFUMI
APPLICANT: OHTA, YASUHIRO
APPLICANT: TSUKADA, YOJI
TITLE OF INVENTION: PRIMERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553, 703A
FILING DATE: 30-NOV-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1209 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-553-703A-5

Alignment Scores:
Pred. No.: 0.0142 Length: 1209
Score: 101.50 Matches: 79
Percent Similarity: 38.97% Conservative: 34
Best Local Similarity: 27.24% Mismatches: 98
Query Match: 5.67% Indels: 79
DB: 1 Gaps: 16

US-09-748-451-2 (1-338) x US-08-553-703A-5 (1-1209)
QY 31 ArgLysGlyCysLeuArgPheGlnLeuProGlnArgLysArg-----LeuGly 47
DB 63 CGTATGCGCTTCTGCGTGGAGCAGCTCCGATCGGAGACAGGCGGCTTCTTCAAGCTG 122
QY 47 sleuTyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerValProAspAsnAl 67
DB 123 CCTGGCGCGGAGCGG-----CGGCTGATACGACACT 155

```

**Tue May 27 09:31:56 2003**

us-09-748-451-2.rni

Page 3

[illegible]

```

? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq Version 1.5
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/553,703A
? FILING DATE: 30-NOV-1995
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Altman, Daniel E
? REGISTRATION NUMBER: 34,115
? REFERENCE/DOCKET NUMBER:
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 714-760-0404
? TELEFAX: 714-760-9502
?
? INFORMATION FOR SEQ ID NO: 9:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1209 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: CDNA
? FEATURE:
? NAME/KEY: Coding Sequence
? LOCATION: 1..1206
? OTHER INFORMATION:
?
US-08-553-703A-9

Alignment Scores:
Pred. No.: 0.0142 Length: 1209
Score: 101.50 Matches: 79
Percent Similarity: 38.97% Conservative: 34
Best Local Similarity: 27.24% Mismatches: 98
Query Match: 5.67% Indels: 79
Gaps: 16

US-09-748-451-2 (1-338) x US-08-553-703A-9 (1-1209)
QY 31 ArgLysGlyCysLeuArgPheGlnLeuProGluArgLysSerArg-----LeuCY 47
Db 63 CGGTGATGGCTTTCTGCGTGGAGACACTCCCGACATGGAGACACGGGGCGCTTCTACAGTG 122
QY 47 sLeuTYrGluNApGcLYThrGluLeuThrGluAspTYrPheProSerValProAspAsnAl 67
Db 123 CCTGGGCGCGCACGG-----CGGGTGTATAGCAGACT 155
QY 67 aGluLeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGlyTYrValSerAspLysArg 87
Db 156 CAACTACCTCTGCGTGTGAGGGAGCGAGGTGTGATG-----TACTGTGCGCTGTACCG 209
QY 87 gPheLeuSerLAlaPheHisGluProGlnValGlyLeuIleGlnAlaAlaGlnGln----- 105
Db 210 CAAGCTTGAGCGCTTCCACCGCCCTGAG-----CTTCTGATGCGGCTAAAGCAGGGCG 263
QY 106 -----LeuLeuGlyAspGluAlaProGlnArgGlnArgLeuLeuAlaAspLeuLe 123
Db 264 CGAATTTTGTGTGGCCATGCGCAGTGCGACACTCTCGAANAANAATGTGCTTGTGTGCT 323
QY 123 uHis-----AsnValSerGlnAsnIleAlaAlaGlu-----Th 134
Db 324 GACGGGGGAGCGCGCGCCGTCAAGTGTGACGAGCAACATCTTAGTGAATGCTTGTACAC 383
QY 134 rArgAlaGluAspProTrpPheGluGlyLeuGluSerArgPheGlnSerLysSerGI 154
Db 384 CATGCCATGAAACAGCTGTGGAGGCTGACGGCGGAGACCGGTACCAAGCAGGACG-- 441
QY 154 yTYrLeuArgTYrSerCysGluSerArgLleArgSerTYrLeuArgGluValSerSer- 173
Db 442 -----GTGACATGATGATGACATGATGTGCTGACTGGGTGCGAGGACCCCTCTGG 491
QY 174 -----TYrProSerThrValGlyValaGluAlaGlnGlnGlu----- 185

```





REGISTRATION NUMBER: 31,284  
 REFERENCE/DOCKET NUMBER: W0101CIP  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (404)873-8794  
 TELEFAX: (404)873-8795  
 INFORMATION FOR SEQ ID NO: 1:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2729 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: cDNA  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 FRAGMENT TYPE: N-terminal  
 ORIGINAL SOURCE:  
 ORGANISM: Rabbit  
 CELL TYPE: Reticulocytes  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 113..2149  
 OTHER INFORMATION: /note= "Expression of HRI  
 OTHER INFORMATION: mRNA in Human erythroid cells, using as the  
 OTHER INFORMATION: probe rabbit HRI cDNA from nucleotides 113  
 OTHER INFORMATION: to 2149."  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 229..249  
 OTHER INFORMATION: /note= "Primer used in the  
 OTHER INFORMATION: amplification of human HRI cDNA sequence  
 OTHER INFORMATION: using the rabbit HRI cDNA sequence."  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 543..560  
 OTHER INFORMATION: /note= "Primer used in the  
 OTHER INFORMATION: amplification of human HRI cDNA sequence  
 OTHER INFORMATION: using the rabbit HRI cDNA sequence."  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 448..468  
 OTHER INFORMATION: /note= "Primer used in the  
 OTHER INFORMATION: amplification of human HRI cDNA sequence  
 OTHER INFORMATION: using the rabbit HRI cDNA sequence."  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: 1009..1031  
 OTHER INFORMATION: /note= "Primer used in the  
 OTHER INFORMATION: amplification of a human HRI cDNA sequence  
 OTHER INFORMATION: using the rabbit HRI cDNA sequence."  
 PUBLICATION INFORMATION:  
 AUTHORS: Chen, Jane J.  
 AUTHORS: London, Irving M.  
 TITLE: Cloning of the cDNA of the heme-  
 TITLE: regulated eukaryotic initiation factor  
 TITLE: 2alpha (elf-2alpha) kinase of rabbit reticulocytes:  
 TITLE: Homology to yeast GCN2 protein kinase and human  
 TITLE: double-stranded-RNA-dependent  
 JOURNAL: Proc. Natl. Acad. Sci. U.S.A.  
 VOLUME: 88  
 PAGES: 7729-7733  
 DATE: September-1991  
 US-07-938-782A-1

Alignment Scores:  
 Pred. No.: 0.247 Length: 2729  
 Score: 96.00 Matches: 62  
 Percent Similarity: 35.43% Conservative: 45  
 Best Local Similarity: 20.53% Mismatches: 121  
 Query Match: 5.37% Indels: 74  
 DB: 1 Gaps: 13

US-09-748-451-2 (1-338) x US-07-938-782A-1 (1-2729)

QY 4 LysProLysSerValLysLeuAlaLysSerProArgLysPheGlyValAlaGly 23  
 DB 1427 AAGCTAGAAATATTTCTCATGCTCTGATCAACAAGTAAAGAAATAGAGACTTGTG 1486  
 QY 24 ArgSerGySGInGluValLeuArgLysGlyCysLeuArgPhe 38  
 DB 1487 CTGGCTGGCGGCATATCATCAAGAAATGCGGCCGACACAGAAAGGGGAGAGA 1546  
 QY 39 LeuProGluArgGlySerArgLeu-----CysLeuTyrgLysArgLysThr 56  
 DB 1547 GCACCCACACACACTTCCCGAGTGGGACCGTGTGTACGCTCGCCGACCAATTGGAA 1606  
 QY 57 GluAspTyrgPheProSerValProAsp-----AsnAlaGluValLeuThrLeu 74  
 DB 1607 GGATCGGAGTATGATGCCAAGTACAGATGTACAGCTCGCGGATCTGTGAGACTG 1666  
 QY 75 GlyAlaIleArgPheGlnGlyValSerAspLeuArgLysPheLeuSerAlaPheHisGlu 94  
 DB 1667 -----TTCCAGCCCTTCGGGACAGATGGAGCGG--GCAGAGTCTGACGGGC 1714  
 QY 95 ProGluValGlyLeuIleGlnAlaIleGlnLeuLeuCysAspGluAlaProGln 114  
 DB 1715 GTGGAGCTGGCGGCATATCCGAGCTCCCTGAGTAAAGAGTCCCGGCGGACCAATAC 1774  
 QY 115 ArgGlnArgLeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaIleGluThr 134  
 DB 1775 GTCCAGCTGCTGACACAG-----AGGAACCGCTCCAGCGG----- 1810  
 QY 135 ArgAlaGluAspProPheGlnGlyLeuGlnSerArgPheGlnSerGly 154  
 DB 1811 -----CGTCCGCGCTTCAGCTGTGAGAGTGAAGTCTTCCAGAACTCCGCG 1858  
 QY 155 TyrLeuArgTyrgSerGlySerArgIleArgSerTyrgLeuArgLysSerTyrg 174  
 DB 1859 CATGTAAACGTACACCTACAGATGATAGATAGAGAGAAAGAAATCGAGACTC 1918  
 QY 175 ProSerThrValGlyAlaGlnGluGlnGluPheLeuArg----- 188  
 DB 1919 AAGAACAGCTGAGCTCTCTCCAGGCGGAGGAGTGAAGGAGGAGAGCGA 1978  
 QY 189 -----ValLeuGlySerMetCysGlnArgLeuArg 198  
 DB 1979 GAGTCCCTGCTAGCCGTACTCGGCCAGCTACAGGAAAGCTGACCTTGCACTGCA 2038  
 QY 199 SerMetGlnTyrgAsnGlySerTyrgPheAspArgGlyAlaLysGlyGly----- 214  
 DB 2039 GCACTCAACAGGATGACATTTCAAGCCCTCGAGGTTGACGGCGCAATATCTGACT 2098  
 QY 215 -----SerArgLeuCysThrProGlnGlyTyrgPheSerCysGlnGlyProPhe 230  
 DB 2099 TGAATCACTCAGCCCATGATGACTCTCC-----TCATCC----- 2134  
 QY 231 AspMetAspSerCysLeuSerArgHisSerIleAsnProTyrgSerAsnArgGluSerArg 250  
 DB 2135 -----TCCTTCTCCGAGG-----TAC 2152  
 QY 251 IleLeuPheSerThrTyrgAsnLeuAspHisIleIleGluLysLysArgThrIleLeuPro 270  
 DB 2153 CTCCTGTGACCTCTGCTGACTG-----CTCCCAATTAACTTACGCTTTTCCCT 2203  
 QY 271 ThrIleu 272  
 DB 2204 TTCCTA 2209

RESULT 7  
 US-08-630-524-1  
 Sequence 1, Application US/08630524  
 Patent No. 5690930  
 GENERAL INFORMATION:  
 APPLICANT: Chen, Jane J.  
 APPLICANT: London, Irving M.  
 TITLE OF INVENTION: DNA Encoding the Heme-Regulated  
 Eukaryotic Initiation Factor 2 alpha kinase







14 ArgSerProArglyspheglnValAlaGlyArgSerCysglnGluValLeuArglyGly 33  
253 CGGACAGTCCGGAAA---GGACTGACAGCTGCACCTCCAGAACTGATGATG--- 306  
34 CysLeuArgpHeGlnLeuProGluArgGlySerArgCysLeuArglyGly 53  
307 GTCTTGAGACCTTGCTGCTACGTGAGTCTTAACACTGCTCTGGAGAGGAGGACT 366  
54 GluLeu---ThrglnAspTyrPheProSerValProAspAsnAlaGluLeuValLeu 72  
367 GCTGTGACAGTGAAGACTCTTCCACGCTGTGAGACGACGACGCTGATGCTCT 426  
73 ThrLeuGlyGlnAlaTrp-----GlnGlyTyrValSer----- 83  
427 GAGCAGGGCCAGAGCTGAGCTGAGTCTTAACACTGCTCTGGAGAGGAGGACT 486  
84 -----AspLeuArgArgPheLeuSerAlaPheHisGluProGln--- 96  
487 AAGCCAAACACAGCAAGAGACATGCCCCGATCACCTTCGATGTGTCAAGCAAAATCCC 546  
97 -----ValGlyLeuLeuGlnAlaGlnGlnLeu 107  
547 CGAGACCTCTTGGACAGCTCAACGTGAAGCAAACTTATGAGCTCTACTCATGAGC 606  
108 CysAspGlnGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuHisAsnValSer 127  
607 TGTGATTTCCAGAGAGTGGCCCTCAAAAGATCTCAGGAGAGCTCCCTGGGACCTTCC 666  
128 -----GlnAsnLeu-AlaAlaGlnUthrArgAlaGluAspProPro-----Tr 141  
667 TGCAGCTGCAAGGCTGGGCGCATATGCTCTGGGACATCTTCACACCTTCCGCAATGTG 726  
141 pPheGlnGlyLeu 145  
727 GTGAGAGGGGCTG 739

Db

QY

RESULT 11  
US-09-085-199B-8  
Sequence 8, Application US/09085199B  
Patent No. 6235879  
GENERAL INFORMATION:  
APPLICANT: Hayden, Michael R.  
APPLICANT: Hugen, A.H.M. Mahbul  
APPLICANT: Chopra, Vikramjit Singh  
APPLICANT: Kalchman, Michael  
TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
TITLE OF INVENTION: Huntington's Disease Gene  
NUMBER OF SEQUENCES: 44  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Opedahl & Larson  
STREET: PO Box 5270  
CITY: Frisco  
STATE: CO  
COUNTRY: USA  
ZIP: 80443-5270  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage  
OPERATING SYSTEM: MS DOS 5.0  
SOFTWARE: MordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/085,199B  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Larson, Marina T.  
REGISTRATION NUMBER: 32038  
REFERENCE/DOCKET NUMBER: UBC-P-013US2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (970) 668-2050  
TELEFAX: (970) 668-2052  
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2301  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: no  
ANTI-SENSE: no  
ORIGINAL SOURCE:  
ORGANISM: mouse  
FEATURE:  
OTHER INFORMATION: cDNA for Huntington-interacting protein - mHIP1  
US-09-085-199B-8

Alignment Scores:  
Pred. No.: 1.17 Length: 2301  
Score: 89.50 Matches: 69  
Percent Similarity: 36.89% Conservative: 59  
Best Local Similarity: 19.88% Mismatches: 117  
Query Match: 5.00% Indels: 102  
DB: 4 Caps: 15

US-09-748-451-2 (1-338) x US-09-085-199B-8 (1-2301)

QY 22 AlaGlyArgSerCysglnGluValLeuArglyGlyCysLeuArgpHeGlnLeuProGln 41  
506 GACAGAGCGAGCTGAGTGAATGAAGAAAGCCAGGCTAATGAACAG----- 556  
42 ArgGlySerArgLeuGlyCysLeuTyrGlnAspGlyThrGluLeuThrGluAspTyrPhePro 61  
557 AGGTATGACAG-----TTAAAGAGAGACTACGATGCTGTGTCGAGAAACATGTCGAC 610  
62 SerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTrpGlnGlyTyr 81  
611 CTGCTGCGGAAGAACGACAG----- 631  
82 ValSerAspLeuArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuLeuGln 101  
632 -----GTGACCAAAACAGGTGTCGTCGCGCCGCAACCCGAGTGTGGAAGA 682  
102 AlaAlaGlnGlnLeuLeu-----CysAspGlnGlnAlaProGlnArgGln 116  
683 GAGAAAAAGAGCTAGCAGATTCTTGCACGTGTAAGTACACGCGCCAGCGAGAGACT 742  
117 ArgLeuLeuAlaAspLeuLeuHisAsnValSerGlnAsnLeuAlaAlaGlnUthrArgAla 136  
743 CAAGAGCAACAGATGTTTACAGAACCTGAGCATGACCTGGCCACC---AGCAGACAG 799  
137 GluAspProProTrpPheGlnGlyLeuGlnSerArgPheGlnSerTyrSerGlyTyrLeu 156  
800 GAGCTGACAGTCTCTCCACAGCAACCTGGAACCTCTCCACAGTACAGAAATGAGCTG 859  
157 ArgTyrSerCysglnGlnSerArgGlyLeuArgSerTyrLeuArgGluValSerTyrProSer 176  
860 ACACAGATCGCCGAG-----TTGAGAGAGAACAGCAGCAGCTTGCG 901  
177 ThrValGlyAlaGlnAlaGlnGlnGluPhe----- 186  
902 ACTGTGACACTGAGAGAGAAAGTATTCAGCCCTCCGAGACACGACGGAAGACAC 961  
187 -----LeuArgValLeuGly-----SerMetCysGlnArgLeuArgSerMetGlnTyr 202  
962 CAGATCAAGCTGGCTGGGCCAGCAAGATTCATGCGCAGAGTGAAGACAGACAGAG--- 1018  
203 AsnGlySerTyrPheAspArgGlyAlaValGlyGlySerArgLeuGlyThrProGlnGly 222  
1019 -----AAACCTCTTGGCAGGAGTCAAGAGCTGCGGAGCGT--- 1057  
223 TrpPheSerCysglnGlyProPheAspMetAspSerCysLeuSerArgHisSerLeuAsn 242  
1058 -----GAGATACAGAGGCGCTGAGCCAG----- 1081  
243 ProTyrSerAsnArgGlnSerArgGlyLeuPheSerThrTrpAsnLeuAspHisLeuLe 262

Db 1082 -----CTTGAGGAAACCCCTCATCAGCTTCGACAGATCCACAGATCCCTTC 1132  
 QY 263 GltysLysArgThrIleIleProThrLeuValGluAlaIleLysGluGlnAspGlyArg 282  
 Db 1133 TCCAAAGTCAGCTCCGCTTCAGCTCCGCTCCGACCACTG-----GAAAGAGAGCGGAGC 1186  
 QY 283 GluValAspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLysLeuVal 302  
 Db 1187 CAG----- 1189  
 QY 303 HisIleValCysHisLys-----LysThrThrHisLysLeuAsn-CysAspPro-- 318  
 Db 1190 TATCTGGCTGCCCGCAAGATATTAGTACGACTTCCTGCATCATCACTCTGCTCCAC 1249  
 QY 319 -----SerAglIleTyrLysProGlnThrArgLeuLysAlaGlySerG 332  
 Db 1250 TTGACCGGTGACACTGTCATCAGGAGGAGTGCACCAAGCCCTCCGGGCCCCACCGAGCCA 1309  
 QY 332 nProValArgLysArgGln 338  
 Db 1310 GCCGACTCTGACGAGG 1328

RESULT 12  
 US-09-061-702-3  
 ; Sequence No. Application US/09061702  
 ; Patent No. 6165737  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Xiaodong  
 ; APPLICANT: Liu, Xuesong  
 ; TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN  
 ; NUMBER OF SEQUENCES: 21  
 ; ADDRESS: Arnold, White & Durkee  
 ; STREET: P. O. Box 4433  
 ; CITY: Houston  
 ; STATE: TX  
 ; COUNTRY: USA  
 ; ZIP: 77210-4433  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: floppy disk  
 ; OPERATING SYSTEM: IBM PC compatible  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/061.702  
 ; FILING DATE: Concurrently Herewith  
 ; CLASSIFICATION:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: McMillian, Nabeha R.  
 ; REGISTRATION NUMBER: P-43,363  
 ; REFERENCE/DOCKET NUMBER: UTSD:546  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (512)418-3000  
 ; TELEFAX: (512)474-7577  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 1689 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: 113..1105  
 ; US-09-061-702-3

Alignment Scores:  
 Pred. No.: 0.814 Length: 1689  
 Score: 89.00 Matches: 78  
 Percent Similarity: 36.18% Conservative: 49  
 Best Local Similarity: 22.22% Mismatches: 146  
 Query Match: 4.97% Indels: 78

DB: 4 Gaps: 15  
 US-09-748-451-2 (1-338) x US-09-061-702-3 (1-1689)  
 QY 5 ProLysSerValLysLeuArgAlaLeuArg-----SerPro 16  
 Db 140 CCAAGATCTGCGGAGATCCGAGCTCTAAAGCCGTCTGTGCGCCCGCAACACAGCCGCC 199  
 QY 17 ArgLysPheGlyValAlaGlyLysSerGlyGluValLeuArg-----LysGlyys-- 34  
 Db 200 GAACACAGCGCGCTGCCGCC-----TCTGCTCGAAGACTAGAGACAGAGCCCTGTAC 256  
 QY 35 ----LeuArgPheGlnLeuProGlnArgGlySerArgLysLeuTyrGlnAspGlyThr 53  
 Db 257 ATTCTGCCATTGATTAAGTCCCTGCACACCACTACACCTTGTCTGCAGAGAGATGACCC 316  
 QY 54 GluLeu----ThrGluAspTyrPheProSerValProAspAsnAlaLeuValLeuLeu 72  
 Db 317 ATAGTGATGATGACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 376  
 QY 73 ThrLeuGlyGlnAlaTrpGlnGlyTyrValSerAspIleArgPheLeuSerAlaPhe 92  
 Db 377 GCTAGTAATGAGAAATG-----GCATAC 400  
 QY 93 HisGluProGlnValGlyLeuIleGlnAlaAlaGlnLeuLeuLysAspGlnAla 112  
 Db 401 AACAAATTCAGATGAGATGACAGCTGATTTCCCAAGATCTTGTATGATAGATCAACA 460  
 QY 113 ProGlnArgGlnArgLeuAsnLeuAsnLeuHisAsnValSerGlnAsnIleAla 132  
 Db 461 GACACGGGGGAGGGGTG-----AAGTGAAGAAATGGCCAGG 499  
 QY 133 GluThrArgAlaGlnAspProProTrpPheGlnGlyLeuGlnSerArgPheGlnSer 152  
 Db 500 GAGCTGAAA-----GAAATCTGCCAGCATCTCTCTTANCA 538  
 QY 153 SerGlyTyrLeuArgTyrSerCysLeuSerArgIleArgSerTyrLeuArgGlnValSer 172  
 Db 539 GAGGAGACCTCAGATCTGTTGTTGATCCCTGCTCAGACCTGCTGAGAACTAGT 598  
 QY 173 SerTyrProSerThrValGlyAlaGlnAlaGlnLeuLeuArgValLysSer 192  
 Db 599 CAGAGTTGGCCACCGCTC---CAGCGCTGAGACACACACTCCAGAGGCTTGACCAA 655  
 QY 193 MetCysGlnArgLeuArgSerMetGlnTyrAsnLysSerTyrPheAspArgGlyAlaLys 212  
 Db 656 AGAGAGAAAGTGGTCCAGTCCAAAGAGCTCCTGAGCTGATCCAGGCTTTGGAGAAA 715  
 QY 213 GlyGlySerArgLeuCysThrProGlnGlyTyrPheSerCysGlnGlyProPheAsp-- 231  
 Db 716 GAGGAGACCTCTTGCACAAAGCAGAA-----GAGTCCAAAGCTGCTTGTGAG 766  
 QY 232 -----MetAspSerCysLeuSerArgHisSerIleAsnProTyrSerAsnArg 247  
 Db 767 GAGGTGATGACGATGACAGCGGATGACGAGAGACC----- 805  
 QY 248 GluSerArgIleLeuPheSerThrTyrAsnLeuAspIleIleGluLysLysArgThr 267  
 Db 806 TCCTCGAGCTTCGCTGGCAGC-----CACATCTT----- 838  
 QY 268 IleIleProThrLeuValGluAlaIleLysGlnAspGlyArgGluValAspTrpGlu 287  
 Db 839 -----ACTGCACTGAGGAGAGAGAGGCTCCAGAG----- 868  
 QY 288 TyrPheTyrGlyLeuLeuPheThrSerGluAsnLeuLysLeuValHisIleValCysHis 307  
 Db 869 -----CTGAGCTTACTAGTACGATTTGAGTTGGTTGCCAAGAGAACGCC 916  
 QY 308 LysLysThrThrHisLysLeuAsnCysAspProSerArgIleTyrLysProGlnThrArg 327  
 Db 917 AAAGCACTGCTGCTGCTTGAACGAGACATAAAGAGAGAGAGAGAGAGAGAGAGAGAGCC 976  
 QY 328 LeuLysArgLysGlnProValArgLysArgGln 338

Db 977 TGTAGCGGAGCTGCCCTGCGCTGACAG 1009

## RESULT 13

5223423-1

Patent No. 5223423

APPLICANT: FRANCHINI, GENOVEPPA, MONG-STAL, FLOSSIE;

TITLE OF INVENTION: CHARACTERIZATION OF REPLICATION COMPETENT

HUMAN IMMUNODEFICIENCY TYPE 2 PROVIRAL CLONE HIV-2 SBL/ISY

NUMBER OF SEQUENCES: 4

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/331,212

FILING DATE: 03-31-1989

SEQ ID NO: 1

LENGTH: 9633

## Alignment Scores:

No.	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
15.6	9633	40	20	55	29	7
88.50						
41.67%						
Local Similarity:						
4.95%						

US-09-748-451-2 (1-338) x 5223423-1 (1-9633)

16 ProctarglyspheglyValAlaGlyArgSerCysGlnGluValLeuArgGlyGlyLeu 35  
 2658 CCACACCCAGCAGGATTAACCAAGAAAGAAATTAAGTGTAGAGATGCC 2717  
 Db 2718 TACTTTCATACCC-----CTATATGAGATTTAGACAGAT 2756  
 Oy 56 ThrgluaspyrphrproserValProaspasnaIaGlu-----LeuValLeu 71  
 Db 2157 ACTGATTTACTGTCATCAGTA--AACATGCAAGCAAGAAAGATATATATAC 2813  
 Oy 72 LeuthrleuglyGlnAlaTrpGlnGly-----TyrValSerAspIleArgArg 87  
 Db 2814 AAGGTCTTACCAAGGATGAGAGGATGACCAATTTTCAATACATAGAGCA 2873  
 Oy 88 PheleuSerAlaPheHISglu-----ProGlnValGlyLeuIleGlnAlaGln 105  
 Db 2874 GCTTATAGAACATCAAGAAAGCAAGACCATGTCATTATGTCAGTACATGATGAT 2933  
 Oy 106 LeuLeuGlyAspGlnAlaTrpGlnArgGlnArgLeu----- 118  
 Db 2934 ATCTTAATA-----GCTAGTACAGACGACATTTGGAACATGACAAAGTACCTTA 2984  
 Oy 119 ----LeuAlaAspLeuHISAsnValSerGlnAsnIleAlaGlnIleThrArgAlaGlu 137  
 Db 2985 CAGGTAAAGAACTTCATTAATGGCTGGATTTCCACCCAGCAAGAAAGTCCAAAG 3044  
 Oy 138 AspProProTrp 141  
 Db 3045 GACCTCCATAC 3056

## RESULT 14

US-09-069-023-13

Sequence 13, Application US/09069023A

Patent No. 6348573

GENERAL INFORMATION:

APPLICANT: Nunez, Gabriel

APPLICANT: Inohara, Naohiro

APPLICANT: Koseki, Takeyoshi

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS

FILE REFERENCE: UM-03333

CURRENT APPLICATION NUMBER: US/09/069,023A

CURRENT FILING DATE: 1998-04-27

NUMBER OF SEQ ID NOS: 38

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 13

LENGTH: 1049

TYPE: DNA

ORGANISM: Homo sapiens

US-09-069-023-13

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.5	1049	61	31	85	37	10
88.00						
43.19%						
Best Local Similarity:						
28.64%						
Query Match:						
4.92%						

US-09-748-451-2 (1-338) x US-09-069-023-13 (1-1049)

14 ArgSerProArglyspheglyValAlaGlyArgSerCysGlnGluValLeuArgGlyGly 33  
 Db 195 AGGACGACGCGCGCT---GGGATGATGCAAGCAGCCTGCAAGGATCATCAGCAAGCT 251  
 Oy 34 CysleuArgpheGlnLeuProGlnArgGlySerArgLeuGlyGlyGlyGlyGlyGly 53  
 Db 252 CTGATGCGCTGCTATCGCTACCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 311  
 Oy 54 GluLeu---ThrgluaspyrphrproserValProaspasnaIaGluLeuValLeu 72  
 Db 312 GTGCTGACACAGAAAGATTTTCAACACCTTGGAGACCAACAGCAATTCATCATCTTG 371  
 Oy 73 ThrlleuglyGlnAlaTrpGlnGlyTyrValSerAspIleArgArgPheLeuSerAlaPhe 92  
 Db 372 GAAAGAGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 430  
 Oy 93 His-----GluProGlnValGlyLeuIleGlnAlaGlnIleGlnLeuGlyGly 108  
 Db 431 GAGTGGGATTAACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 487  
 Oy 109 AspGlnGlnAlaProGlnArgGln-ArgLeuLeuAlaAspLeuHISAsnValSerGln 128  
 Db 488 ---CATGCGCTTAACTTAACGAGGACCATGATGAG--ATGATCTCGCTGCTTA 541  
 Oy 128 nasIleAlaAlaGluThrArgAlaGluAspProProTrpPheGlnGlyGlyGlyGly 148  
 Db 542 CGACATCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571  
 Oy 148 gPheGlnSerIleSerGlyTyrLeuArgTyrSerCysGlnSerArgIleArgSerTyrLe 168  
 Db 572 GCTGAGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 628  
 Oy 168 uArgGluValSerSerTyrProSerThrValGlyAlaGluAlaGlnGlnGlnGlnGlnGln 188  
 Db 629 CATCATCTGCGGCAATAC-----ARGCTCCG 655  
 Oy 188 gValleuGlySerMetCysGlnArg-----LeuArgSerMetClnTyrAsn-GlySerT 206  
 Db 656 GGTCTGATGACAGAGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 715  
 Oy 206 YrPheAspArgGlyAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 217  
 Db 716 GTGTGATAGGATGCA---GGCTGTGCGCGGCTC 747

## RESULT 15

US-08-794-158-1

Sequence 1, Application US/08794158

Patent No. 6387655

GENERAL INFORMATION:

APPLICANT: Burnett Jr., J. Paul

APPLICANT: Mayne, Nancy G.

APPLICANT: Sharp, Robert L.

TITLE OF INVENTION: Excitatory Amino Acid Receptor Protein

NUMBER OF SEQUENCES: 3



TELEPHONE: 317-276-3334  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2637 base pairs  
 TYPE: nucleic acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: mRNA  
 HYPOHETICAL: NO  
 ANTI-SENSE: NO  
 US-08-794-158-3

## Alignment Scores:

Pred. No.:	2.21	Length:	2637
Score:	88.00	Matches:	75
Percent Similarity:	34.81%	Conservative:	43
Best Local Similarity:	22.12%	Mismatches:	113
Query Match:	4.92%	Indels:	108
DB:	4	Gaps:	17

-748-451-2 (1-338) x US-08-794-158-3 (1-2637)

```

46 Leucylserineylglutylthreonylthreonylprolineasp 65
   :::::
374 AUAUGUGUCUGAUGAUCUUAUCCAUCAAGAAACAUCCAUUCUUCAGAGG 433
   :::::
66 AsnAlaGluValLeuLeuThrLeuGlyAlaAlaTrpGlnGlyTyrValSerAsp 85
   :::::
434 UCAUGUGUGUCUUAUAGAGUGUUUCCAUACAGUGGCAACUUCGCGUCUUC 493
   :::::
86 ArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAla-AlaGln 105
   :::::
494 AGAUGCCUUCAGACUUCAGUACGACGACGACCAACUACGAGUAAUAGCGCUG 553
   :::::
105 nLeuLeuCyAspGluGlnAlaTrpGlnAlaArgGlnArgLeuAlaAspLeuHis-- 124
   :::::
554 AUUACUUCUCCAGAGACGUGUCCGCCCC-----CGACUUCUACAGCCCAAGCC 604
   :::::
125 -----AsnValSerGlnAsnIleAlaAla----- 132
   :::::
605 CUGAGAUUCUGCGCUCUUCACUGACGACGUCGUCACAGAGAGCCUGAGGUGAU 664
   :::::
133 -----GluThrArgAlaGluAspProTrpPheGluGlyLeuG 146
   :::::
665 ACGGGAGACAGGAGUACGAGCCUUCUAGAGAGAGCCGCU---CGCAACAUUCGA 721
   :::::
146 uSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerArgIleArg 166
   :::::
722 U---CGUACGGGGAGAGAGUGGGCGCCCAACAUCCCAAGUCCUACAGAGCGUGA 778
   :::::
166 rTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGlnAlaGln----- 183
   :::::
779 U-----CCGAGAACUGUU---GCAAGAACCCCAAGCGCGCG 811
   :::::
184 -----GluGluPheLeuArgValLeuGlySerMetCysGlnArg-----Le 197
   :::::
812 UCGUGGUCUUCUACAGCGAGCGAGCAGCGCGGAGUCCAUUGCAGCGCGCGCG 871
   :::::
197 uArgSerMetGlnTyrAsnGlySerTyrPheAspArgAlaAlaGly----- 213
   :::::
872 CCAUAGCUCUUCUACUUGGUGGCCAGCAGCGCGUGGGCGCGAGAGAGAUCAUCA 931
   :::::
214 -GlySerArgLeuCys-----ThrProGluGlyTrpPheSerCysGlnG 228
   :::::
932 AGGAGCAGAGCAUAGUGGCGUACGAGCGCAUACCCUGAGUGGCUCCAGCGUGUC 991
   :::::
228 yPro-PheAspMetAspSerCysLeuSerArgHis-----SerIleAspProTyrSera 246
   :::::
992 GCCAGUUCGAC-----CGCUACUUCAGAGACCCUACACCCUACACACA 1033
   :::::
246 sn----- 246
   :::::
1034 ACCACCGCAACCCUUGUGCGGAGACUUCUGGAGCAAAAGUUUUCAGUCCAGCCUCCAGA 1093

```

```

QY 246 ----- 246
DB 1094 ACAAGCAACACAGCGGUGUGGACAGACACUUGCCAUACAGACAGCAUACG 1153
QY 247 --ArgGluSerArgIleLeuPhe-----SerThrTrpAsnLeuAspHisIleIleG 263
DB 1154 AGCAAGAGUCCAAAUAGUUGUGUGGAGAACGCGUGUAGCCAUUGCCAGCGUG 1213
QY 263 lLys--LysArgThrIleIleProThr-----LeuValGluAlaIleTyrGluG 279
DB 1214 ACAAAUAGCAGCGACCCUUCUCCCAACACUACCAAGCUUUGAGUCCAUAGAACUCC 1273
QY 279 lAspGlyArgGluValAlaAspTrpGluTyrPheTyrGlyLeuLeuPheThrSer 296
DB 1274 UGGAUGGGAAGAAAGUUGUACAGAGAUUACUUGCGAUAUACACUUCACGCGCU 1326

RESULT 17
US-08-537-342-1
: Sequence 1, Application US/08537342
: Patent No. 5840486
: GENERAL INFORMATION:
: APPLICANT: Pederesen, Oluf
: APPLICANT: Bjorbaek, Christian
: TITLE OF INVENTION: Mutant DNA Encoding Protein Phosphatase
: TITLE OF INVENTION: 1 G-Subunit
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: No. 5840486 No. 5840486disk of No. 5840486th America, Inc.
: STREET: 405 Lexington Avenue, 64th Floor
: CITY: New York
: STATE: New York
: COUNTRY: United States of America
: ZIP: 10174-6401
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/537,342
: FILING DATE: 02-OCT-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Agtis, Dr. Cheryl H.
: REGISTRATION NUMBER: 34,086
: REFERENCE/DOCKET NUMBER: 4276, 050-US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-867-0123
: TELEFAX: 212-878-9655
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4322 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-537-342-1

Alignment Scores:
Pred. No.: 7.49 Length: 4322
Score: 86.50 Matches: 84
Percent Similarity: 33.41% Conservative: 53
Best Local Similarity: 20.49% Mismatches: 162
Query Match: 4.84% Indels: 111
DB: 2 Gaps: 18

US-09-748-451-2 (1-338) x US-08-537-342-1 (1-4322)
QY 5 ProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGlyArg 24
DB 1364 CCCTCTTCACATCACTAATGTCGAGGAACCTTAATAAACAAT---GAAGAGAGAGCT 1420

```

```
OY 25 SerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuProGlnArgLysSer 44
DB 1421 AAAAAATTGAAGTAAAGATTGGATGTTTACGAAAGATTTC---CATTCAGATATAC 1477
OY 45 ArgLeuCysLeuTyrGluAspGlyThrGlu-----LeuThrGluAspTyrPheProSer 62
DB 1478 TCGCATGCTCTCAAGATCAAGAAAGAGATCTTCTCAAGAGATGATATATATGCGCAAT 1537
OY 63 ValProAspAsnAlaGlu---LeuValLeuLeuThrLeuGlyGlnAlaIleProGlnTyr 81
DB 1538 GGTAAAGATGATGATGAAACAAAGAAATATATATGATGTTATGATGTTAAACAAAGA----- 1591
OY 82 ValSerAspIleArgPheLeuSerAlaPheHisGlu----- 94
DB 1592 -----AAAATTCCAAATCTTACATGACCAAGAAAGAAAGATGGGTAC 1639
OY 95 ProGlnValGlyLeu-----IleGlnAlaAlaGlnGln-----LeuLeuGly 108
DB 1640 CTTAAATATAGTGTGCGACGAGATTGAGCTAGTACAGACAGACCTGCTACTCTGCTGAC 1699
OY 109 AspGlnGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuLeuHisAsnValSerGln 128
DB 1700 GAACATACCGCATCCCAACCCGGGACATCAAGCAGATGTGCTCATTCACCAAGACACA 1759
OY 129 AsnIleAlaIleGluThrArgAlaGluAspProTyrPheGlnGlyLeuGlnSerArg 148
DB 1760 AATTAAAGTTGGAA---GAAGCTGTGTTAACCCAGACATCATTCATTGACATGTGAA 1816
OY 149 PheGlnSerLysSerGlyTyrLeuArgTyrSerCysGlnSerArgIle----- 164
DB 1817 GGCAGCGCTTATGAGGAGATTAACGTGTCATAGTTGTTATCAAGAACTGGAATGTTTG 1876
OY 165 ---ArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyValGlnAlaGln 183
DB 1877 AGGAATGATTTACTCTTCCAAAGTGAAGAAATAACAGTGGGATTAATTCTGAAATCAG 1936
OY 184 GlnGlnPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetCysIlePyrAsn 203
DB 1937 GAT----- 1997
OY 204 GlySerTyrPheAspArgGlyAlaLysGlyGlySerArg----- 216
DB 1958 CAAAGTTGGAATGTTCTGGAAGAGTACAGGAAATCAAGAGATTAAGCAACATATACA 2017
OY 216 ----- 216
DB 2018 GAGCATATCAAGGCAAAACAGATTGCAAGACGTGGGAAAAAGACATATATCCAGG 2077
OY 217 -----LeuCysThrProGlnGlyTyrPheSerCysGlnGlyProPheAspMetLysP 233
DB 2078 AGTTGAAAGTACTACTACAGAAATGTTTACTGCGCAAGAAACAGTGTCTGTGAAGTG 2137
OY 234 SerCysLeuSerArgHisSerIle----- 241
DB 2138 TCTTCTCTAGCTGATCAGATGCGATTAAGAGAAAGACAGAACTGTATACCTATATATT 2197
OY 242 -----AsnProTyrSerAsnArgGlnSerArgIleLeuPhe 253
DB 2198 AAGACATCATCAAGAAAGTACTCCAGAAAGCATGCTGCTGAGAAAGAAACATATATTCT 2257
OY 254 -----SerThrTyrAsnLeuAspHisIleIleGlnLysLysArgThrIleLeu 269
DB 2258 AAGCTACTCAAGAGACAGCAAGAGTACAGGCGCCATCGAGTTAAAGAAACA----- 2311
OY 270 ProThrLeuValGlnAlaIleLysGlnGlnAspGlyValArgValAspTyrPheLys 288
DB 2312 -----GCGTTGATCCACATGAAAGGAGGAAATGATCATTCACATATATACC 2356
OY 289 -----PheTyrGlyLeuLeuPheThrSerCysLeuLeuLysLeuValHis 303
DB 2357 CTTTGCAACGAGATACAGTACGTATCTATCTATGACAAATATTTTGAAGAAATCACACT 2416

OY 304 Ile---ValCysHisLysLysThrHisLysLeuAsnCysAspProSerArgIleTyr 322
DB 2417 TTACGATATTTTGATATGATCTAGATGATGAAATGAGAGAGAAAGAAACATGCTATGTATC 2476
OY 323 LysProGlnThrArgLeuLysArgLysGln 332
DB 2477 AATCTTGAAGAACATCATGACAGGAGGAGAA 2506

RESULT 18
US-09-069-023-17
; Sequence 17, Application US/09069023A
; Patent No. 6348573
; GENERAL INFORMATION:
; APPLICANT: Nunez, Gabriel
; APPLICANT: Inohara, Naohiro
; APPLICANT: Koseki, Takeyoshi
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING APOPTOSIS
; FILE REFERENCE: JN-03333
; CURRENT APPLICATION NUMBER: US/09/069,023A
; CURRENT FILING DATE: 1998-04-27
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 1114
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-069-023-17

Alignment Scores:
Pred. No.: 0.967 Length: 1114
Percent: 86.00 Matches: 65
Score Similarity: 37.55 Conservative: 27
Best Local Similarity: 26.53 Mismatches: 102
Query Match: 4.81% Indels: 52
DB: 4 Gaps: 13

US-09-748-451-2 (1-338) x US-09-069-023-17 (1-1114)
OY 14 ArgSerProArgLysPheGlyValAlaIleArgSerGlnGlnValLeuArgLysGly 33
DB 343 CGAAGTACGCGCGGT---GGGTGATGCGCCAGCAGCTGTCAGAGACTTATTCAGCAAGACT 399
OY 34 CysLeuArgPheGlnLeuProGlnArgGlySerArgLeuCysLeuTyrGlnAspGlyThr 53
DB 400 CTGAGTCTTACTCTATCAACACTGGCTGTACGCTGTGCTGTGAGAGAGAGCGGCACC 459
OY 54 GluLeu---ThrGluAspTyrPheProSerValProAspAsnAlaGlnLeuValLeuLeu 72
DB 460 GTGCTGACACAGAGAGATTTCTCAGACCTTAAGGAGCAACAGCATTTCTATGATCTTG 519
OY 73 ThrLeuGlyGlnAlaIleP-----GlnGlyTyr 81
DB 520 GAAAGGAGACGAAATGACACACCGGATGATGATGATGATGATGATGATGATGATGATGAT 579
OY 82 ValSerAspIleArgArg-----PheLeuSerAlaPheHisGlnPheGlnValGlyLeu 99
DB 580 AATCGGAGATACCAAGTACAGTTCGCA---CCTATACAGCTGATACCCCA---GGACTT 635
OY 100 IleGlnAlaIleGlnLeuLeuLeuLeuCysAspGlnAlaIleProGlnArgGlnArgLeu 119
DB 636 CCGCGCTGCTCA-----TGT-----CAAGCCAGATATGATGATGATGATGATGATGATGAT 680
OY 120 AlaAspLeuLeuHisAsnValSerGln-----AsnIleAlaIleGlnThr 134
DB 661 GGTCTCTACGACATCCATGACACAGGATTAAGAGCCGCTGTAAAGAAATCTCTAGAGTT 740
OY 135 Arg-----AlaGlnAspProTyrPheGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 151
DB 741 TATGCTTATGCTGACAGATATACGAGCACTTCTGATCTA----- 782
OY 152 LysSerGlyTyrLeuArgTyrSerCysGlnSerArgIleArgSerTyrLeuArgGlnVal 171
```

Db 783 ---TGGGGGACATACATGCTCCGAGTACTGGCGATACAGAGA----- 824  
Qy 172 SerSerTyProSerThValGlyAlaGluAlaGlnGluGluPheLeuArgValLeuGly 191  
Db 825 -----GCAGCATCCTCCCAAGACCTAGACCAAGAGCTGTTCATGCTAACAGG 872  
Qy 192 SerMetCysGlnAlaGluArgSerMetGlnTyrAsnGlySerTyPheAsp---ArgGly 210  
Db 873 GCACAGCTACAGAGCGCCAGACCTGCTCTGTATTATAGCTGTGGATGCCAGAGGA 932  
Qy 211 AlAlaGlyGlySerArgLeuGlySerThProGluGlyTyrPheSerCys----- 226  
Db 933 AGGATGGGGGTGTACCATGAGAGGCTGAGTAGAGAGTCTCTGCAAGAAAGAGCG 992  
Qy 227 ---GlnGlyProPhe 230  
Db 993 GCAGAGGGGCGCTTTC 1007  
RESULT 19  
US-09-103-840A-2/c  
Sequence 2, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103.840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 2  
LENGTH: 4403765  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
FEATURE:  
OTHER INFORMATION: CDC 1551  
OTHER INFORMATION: "n" bases at various positions throughout the sequence  
US-09-103-840A-2  
Alignment Scores:  
Pred. No.: 4.86e+05 Length: 4403765  
Score: 85.50 Matches: 67  
Percent Similarity: 36.64% Conservative: 29  
Best Local Similarity: 25.57% Mismatches: 89  
Query Match: 4.78% Indels: 77  
Gaps: 15  
-748-451-2 (1-338) x US-09-103-840A-2 (1-4403765)  
2 LeuGlnLysProLysSerValLysLeuArgAla-----LeuArgSerProArgLys 18  
Db 201968 CTGAATCTGCCAAGATGCTCAGGGTTTCACCAAAATCTCTCGCTGCTCAGCAGG 201909  
Qy 19 PheGlyValAlaGluArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGln 38  
Db 201908 ---TCGGTGAAGAGTCCAGCTGCTCAGACACA-----TGGTTAGGTTTCGGG 201864  
Qy 39 LeuProGluArgGlySerArgLeuGlySerLysGlnAspGlyThrGluLeuThrGluAsp 58  
Db 201863 TTGGTCGTTG-----ATAAGCTTTTGCACCTGGCGCCAGAAAGCGAGATG----- 201819  
Qy 59 TyPheProSerValProAsp-----AsnAlaGluLeuValLeu 71  
Db 201818 ---TTCCCAAGAGGGGCTGACCGCGCGCGCTGTTGATGACGCGCATGAGGCTC 201762  
Qy 72 LeuThrLeuGlyAlaIleArgGlnGlyTyrValSerAspIleArgPheLeuSerAla 91  
Db 201761 TTAGCGTTGACCAAT-----AGCGGTGACACTGCTCA 201729

Qy 92 PheHisGluProGlnVal-----GlyLeuIleGlnAlaAlaGlnLeuLeuCys 108  
Db 201728 CTGGCATCACCCAGAGATCTGCCACCTGTTGGCTGGCTGATAGTGGTATCTGC 201669  
Qy 109 AspGluGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuLeuHisAsnValSerGln 128  
Db 201668 TCGTCGCGCTTCCGAGT-----GTGTCGAG 201642  
Qy 129 AsnIleAlaAlaGluThrArgAla-----Glu 137  
Db 201641 AACTTAGCACCCTGAGAGGCGCGCTCAGTCCGATAGTCTGATCAAGCTCTCC 201582  
Qy 138 AspProProThrPheGluGlyLeu-GluSerArgPheGlnSerLysSerGlyTyrLeuAr 157  
Db 201581 GACAACACATTCACGACACCGCTTACCGCTCGA----- 201548  
Qy 157 gTyrSerCysGluSerArgIleArgSerTyLeuArgGluValSerSerTyProSerTh 177  
Db 201547 ---TGTCCACCGCGAGTGGCGCTTGTGACGTGAGAGAACGCTGTAATCTCG 201495  
Qy 177 rValGly-----AlaGluAlaGlnGluPheLeuArgValLeuGlySerMet 193  
Db 201494 TAGGGGTGTGCTTGTGCCAACCGGCAAGCGCCCGCGGCAACGCTTGGCGCGC 201435  
Qy 193 tCysGlnArgLeuArgSerMetGlnTyrAsnGlySerTyPheAspArgGlyAlaLysG 213  
Db 201434 CGCGCTCGATCTGAGCAGCACTTCTTACCCAGGA---TGTGTGCTGCGGATGGCTAGG 201378  
Qy 213 yGlySerArgLeuSerThProGluGlyTyrPheSerCysGln-GlyProPheAspMet 233  
Db 201377 CGGCTCTCGG---TGCCGA-----TGGTTGTGGTCCGATGAGAACTTACACAGC 201330  
Qy 233 sp 233  
Db 201329 AT 201328  
RESULT 20  
US-09-103-840A-1/c  
Sequence 1, Application US/09103840A  
Patent No. 6294328  
GENERAL INFORMATION:  
APPLICANT: FLEISCHMAN, Robert D.  
APPLICANT: WHITE, Owen R.  
APPLICANT: FRASER, Claire M.  
APPLICANT: VENTER, John C.  
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM  
FILE REFERENCE: 24366-20007.00  
CURRENT APPLICATION NUMBER: US/09/103.840A  
CURRENT FILING DATE: 1998-06-24  
NUMBER OF SEQ ID NOS: 2  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 4411529  
TYPE: DNA  
ORGANISM: Mycobacterium tuberculosis  
OTHER INFORMATION: H37Rv  
US-09-103-840A-1  
Alignment Scores:  
Pred. No.: 4.87e+05 Length: 4411529  
Score: 85.50 Matches: 67  
Percent Similarity: 36.64% Conservative: 29  
Best Local Similarity: 25.57% Mismatches: 89  
Query Match: 4.78% Indels: 77  
Gaps: 15  
US-09-748-451-2 (1-338) x US-09-103-840A-1 (1-4411529)  
Qy 2 LeuGlnLysProLysSerValLysLeuArgAla-----LeuArgSerProArgLys 18  
Db 201801 CTGAATCTGCCAAGATGCTCAGGGTTTCACCAAAATCTCTCGCTGCTCAGCAGG 201742

```
OY      19  PheGlyValAlaIleArgSerCysGlnIleValLeuArgLysGlyCysLeuArgPheGln 38
DB      201741  ---TCGGTAGAGATCCGACGCTGCTCAGACACA-----TGGTCAAGGTTCCGG 201697
OY      39  LeuProGlnArgGlySerArgLeuCysLeuTyrGlnAspGlyThrGlnLeuThrGlnAsp 58
DB      201696  TTGGTCGTTG-----ATTAAGGTTTGGACCTGGGCGGAGAAAGCGAGATG----- 201652
OY      59  TyrPheProSerValProAsp-----AsnAlaGlnLeuValLeu 71
DB      201651  ---TTCCCGACGAGGGCGCTCAGCCGCGCGCTGCTGAGACGCGCGATCAGGCTC 201595
OY      72  LeuThrLeuGlyGlnAlaIleProGlnGlyTyrValSerAspIleArgArgPheLeuSerAla 91
DB      201594  TTAGCGTTGACCAAT-----AGCGCGTGCACCTGCTCA 201562
OY      92  PheHisGlnProGlnVal-----GlyLeuIleGlnAlaIleGlnLeuLeuCys 108
DB      201561  CTGCGATCACCCAGAGATGCTGGCAGCCTGGTTGGCCTAGGCTAGTGGCTGATCTGC 201502
OY      109  AspGlnGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuLeuHisAsnValSerGln 128
DB      201501  TCGTCGCGCTTCCGATG-----GTGTCGAG 201475
OY      129  AsnIleAlaAlaGlnThrArgAla-----Glu 137
DB      201474  AACCTAGCCACCCCGTCGAGGGCGCGCTCAGTGCAGATAGGCTGATCAACGGCTCC 201415
OY      138  AspProProTrpPheGlnGlyLeu-GluSerArgPheGlnSerLysSerGlyTyrLeuArg 157
DB      201414  GACACACATTCAGCGACCGCTGACCTCGA----- 201381
OY      157  gTyrSerCysGlnSerArgIleArgSerTyrLeuArgGlnValSerSerTyrProSerThr 177
DB      201380  -----TGTCCACGACCGGATGCGGCTTGTGTACGTCGACAGACGCGTCGTAAATCTGG 201328
OY      177  rValGly-----AlaGlnAlaGlnGlnGlnIlePheLeuArgValLeuGlySerMe 193
DB      201327  TACGGGGGTGGCTTTGCCCAACCGCAAAACGCCCCCGCGGCAACGCTGGGCGCGCG 201268
OY      193  tCysGlnArgLeuArgSerMetGlnTyrAsnGlySerTyrPheAspArgGlyAlaLysGln 213
DB      201267  CGCGGCTCGATCTCGAGACATTCTCTCCACAGGA--TGGTGTGCGGTGCGGATGGCTAGG 201211
OY      213  xGlySerArgLeuCysThrProGlnGlyTyrPheSerCysGln-GlyProPheAspMetA 233
DB      201210  CGGCTCTCG--TCCCA-----TGGTGTGTTGGTCCGATGAGAACTTGACACAG 201163
OY      233  sp 233
DB      201162  AT 201161
```

Search completed: May 26, 2003, 17:28:03  
Job time : 2100 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM protein - nucleic search, using frame\_plus\_p2n model

Run on: May 26, 2003, 15:26:43 ; Search time 197 Seconds  
(without alignments)  
2265.565 Million cell updates/sec

Title: US-09-748-451-2  
Perfect score: 1789  
Sequence: 1 MLCRKSVMKALMSPKRFG.....SRYYQRTMKRKPVRKQ 338

Mapping table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 828747 seqs, 660231138 residues

Total number of hits satisfying chosen parameters: 1657494

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame+ p2n.model -DEV-xlh  
-O/cg2\_1/USPTO.spool/US09748451/runat 21052003 153830 16347/app.query.fasta.1.519  
-DB-published.Applications.MA -OFM-fastap -SUPPL-rrph -MIRMArch-0.1  
-LOOCC-0 -LOOEXT-0 -UNITS-bits -START-1 -END-1 -MATRIX-blosum62  
-TRANS-human40.cst -LIST-45 -DOCALLIGN-200 -THR.SCORP-0 -THR.MAX-100  
-THR.MIR-0 -ALIGN-20 -MODE-LOCAL -OUTPMT-pto -NORW-ext -HEA5T5E-500 -MINLEN-0  
-MAXLEN-2000000000 -USPR-US09748451.ECGN.1.1.57 6/runat\_21052003\_153830\_16347  
-NCPI-6 -ICPI-3 -NO.MAP -LARGEDUERY -NEG.SCORP-0 -WAIT -DSBLOCK-100  
-IONLOG -DEV.TIMEOUT-120 -MIR.TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5  
-FGAPOP-6 -FGAPEXT-7 -FGAPOP-10 -FGAPEXT-0.5 -DELOP-6 -DELEXT-7

## Database:

Published Applications.MA: \*  
1: /cg2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq: \*  
2: /cg2\_6/ptodata/1/pubpna/PCY\_NEW\_PUB.seq: \*  
3: /cg2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq: \*  
4: /cg2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq: \*  
5: /cg2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq: \*  
6: /cg2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq: \*  
7: /cg2\_6/ptodata/1/pubpna/PCYUS\_PUBCOMB.seq: \*  
8: /cg2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq: \*  
9: /cg2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq: \*  
10: /cg2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq: \*  
11: /cg2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq: \*  
12: /cg2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq: \*  
13: /cg2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq: \*  
14: /cg2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq: \*  
15: /cg2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1789	100.0	2839	10 US-09-748-451-1	Sequence 1, Appl 1
2	102.5	5.7	1766	9 US-10-242-963-11	Sequence 11, Appl 1
3	102.5	5.7	1862	9 US-10-161-803-29	Sequence 29, Appl 1
4	102.5	5.7	1864	9 US-09-974-298-43	Sequence 43, Appl 1

5	102.5	5.7	1892	12 US-10-044-090-793	Sequence 793, App
6	102.5	5.7	3405	9 US-09-974-298-44	Sequence 44, Appl 1
7	102.5	5.7	3405	9 US-09-981-353-36	Sequence 36, Appl 1
8	99.5	5.6	1218	9 US-10-043-487-101	Sequence 101, Appl 1
9	99.5	5.6	2209	9 US-10-161-803-30	Sequence 30, Appl 1
10	99.5	5.6	69327	10 US-09-777-921A-3	Sequence 3, Appl 1
11	97.5	5.4	1749	9 US-09-923-779-149	Sequence 149, App
12	96.5	5.4	2673	9 US-10-060-425-7	Sequence 7, Appl 1
13	95.5	5.3	1922	9 US-09-995-898A-32	Sequence 32, Appl 1
14	95.5	5.2	1225	10 US-09-216-393-284	Sequence 284, App
15	92.5	5.2	1225	10 US-09-216-393-284	Sequence 284, App
16	90	5.0	1282	9 US-09-764-881-988A	Sequence 988A, App
17	90	5.0	1282	9 US-09-764-881-988A	Sequence 988A, App
18	90	5.0	4967	10 US-09-834-975-958	Sequence 958, App
19	90	5.0	11598	9 US-10-091-572-887	Sequence 887, App
20	90	5.0	11598	9 US-09-764-891-9340	Sequence 9340, App
21	90	5.0	11600	9 US-10-091-572-888	Sequence 888, App
22	90	5.0	11600	9 US-09-764-891-9341	Sequence 9341, App
23	89.5	5.0	11770	9 US-10-097-065-121	Sequence 121, App
24	89	5.0	1689	10 US-09-748-451-3	Sequence 3, Appl 1
25	89	5.0	3309400	9 US-09-738-626-1	Sequence 1, Appl 1
26	88.5	4.9	735	10 US-09-770-149-82	Sequence 82, Appl 1
27	88.5	4.9	1074	10 US-09-974-300-2794	Sequence 2794, App
28	88.5	4.9	2084	10 US-09-822-849A-526	Sequence 526, App
29	88	4.9	4235	10 US-09-764-877-3368	Sequence 3368, App
30	88	4.9	10955	10 US-09-070-927A-236	Sequence 236, App
31	87.5	4.9	579	10 US-09-974-300-2935	Sequence 2935, App
32	87.5	4.9	606	10 US-10-165-800-10	Sequence 10, Appl 1
33	87.5	4.9	1284	9 US-10-165-800-12	Sequence 12, Appl 1
34	87	4.9	1728	10 US-09-925-300-555	Sequence 555, App
35	87	4.9	2602	9 US-09-974-298-45	Sequence 45, Appl 1
36	87	4.9	17239	9 US-10-091-504-2326	Sequence 2326, App
37	87	4.9	17239	10 US-09-764-869-2326	Sequence 2326, App
38	86.5	4.8	657	9 US-10-165-800-12	Sequence 12, Appl 1
39	86	4.8	5387	9 US-10-001-873-72	Sequence 72, Appl 1
40	85	4.8	868	9 US-09-938-842A-1748	Sequence 1748, App
41	85.5	4.8	2019	9 US-10-097-340A-17	Sequence 17, Appl 1
42	85.5	4.8	2274	9 US-10-097-340-17	Sequence 17, Appl 1
43	85.5	4.8	14886	9 US-10-203-224-17	Sequence 17, Appl 1
44	85	4.8	353	9 US-09-918-995-32209	Sequence 32209, App
45	85	4.8	1498	9 US-09-800-602-26	Sequence 26, Appl 1

## ALIGNMENTS

RESULT 1  
US-09-748-451-1  
Sequence 1, Application US/09748451  
Patent No. US20010011078A1  
GENERAL INFORMATION:  
APPLICANT: WANG, XIAODONG  
APPLICANT: LIU, XUESONG  
TITLE OF INVENTION: DNA FRAGMENTATION FACTOR INVOLVED IN APOPTOSIS  
FILE REFERENCE: US05:546USD1  
CURRENT FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 09/061,702  
PRIOR FILING DATE: 1998-04-16  
NUMBER OF SEQ ID NOS: 21  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 2839  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-748-451-1

Alignment Scores:  
Pred. No.: 5.83e-223  
Score: 1789.00  
Percent Similarity: 100.00%  
Best Local Similarity: 100.00%  
Query Match: 100.00%  
DB: 10  
Gaps: 0

US-09-748-451-2 (1-338) x US-09-748-451-1 (1-2839)

QY 1 MetLeuGlnYsProLYsSerValIysLeuArgAlaLeuArgSerProArgLYsPheGly 20  
 Db 132 ATGCTCCAGAAAGCCCAAGAGCGTAAAGCTGGGGCCCTGCCAGCCGAGAAATTGGCC 191  
 QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLYsGlyCysLeuArgPheGlnLeuPro 40  
 Db 192 GTGGCTGGCCGAGAGCTGCGAGAGAGTGTGGCAAGGGCTGTCCGCTTCACGCTCCCT 251  
 QY 41 GlnArgGlySerArgLeuGlyLeuTYrGlnAspGlyThrGluLeuThrGluAspTYrPhe 60  
 Db 252 GAGCGGGGTTCGCGCTGTGCTTACGAGGATGCGAGGAGCGAGCGGAAGTAACTTC 311  
 QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaTrpGlnGly 80  
 Db 312 CCCAGTGTCCCGCAACAGCCGAGCTGTGCTGCTCACTTGGGCGAGCGCTGGCAGGGC 371  
 QY 81 TYrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIle 100  
 Db 372 TATGTGAGCGACATCAGAGCGCGCTTCTCAGTCAATTCACGACCAAGTGGGCTCATC 431  
 QY 101 GlnAlaIaGlnGlnLeuLeuCysAspGlnGlnAlaProGlnArgGlnArgLeuAla 120  
 Db 432 CAGGCGCCCGCAGCAGCTGTGTGTAGAGCAGGCGCCACAGAGGAGAGCTGTGCT 491  
 QY 121 AspleuLeuHisAsnValSerGlnAsnIleAlaIaGluThrArgAlaGluAspProPro 140  
 Db 492 GACCTCTGCAACAGTCAGCAGCAGACATCGCGCGCCGAGACCGGGCTGAGAGACCGCG 551  
 QY 141 TrpPheGluGlyLeuGlnSerArgPheGlnSerIysSerGlyTYrLeuArgTYrSerCys 160  
 Db 552 TGGTTTAAAGGCTTGAGTCCCGATTCAGAGCAAGCTGGCTATCTGAGTAAAGCTGT 611  
 QY 161 GluSerArgIleArgSerTYrLeuArgGluValSerSerTYrProSerThrValGlyAla 180  
 Db 612 GAGAGCGCGATCCGAGATTACCTGAGGAGAGTGAAGTCTTACCTCCACAGTGGAGTGG 671  
 QY 181 GlnAlaGlnGlnGlnLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMet 200  
 Db 672 GAGGCTCAGAGAGAAATTCCTGCGGGTCTCGGCTCCATGTCAGAGGCTCGGTCATG 731  
 QY 201 GlnGlyTrpAsnGlySerTYrPheAspArgGlyAlaIaGlyGlySerArgLeuCysThrPro 220  
 Db 732 CAGTACATGCGACCTACTTCCAGACAGAGAGCCAAAGGGCGGACCGCTCTGACACCG 791  
 QY 221 GlnGlyTrpPheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSer 240  
 Db 792 GAAGGCTGTTCTCTCCAGAGGCTCTTTCACATGAGACAGCTCTTATCAAGACACTCC 851  
 QY 241 IleAsnProTYrSerAsnArgGluSerArgIleLeuPheSerThrTrpAsnLeuAspHis 260  
 Db 852 ATCAACCCCTACAGTACAGAGAGAGAGAGATCTCTTACAGACACTGGAACTGGATCAC 911  
 QY 261 IleIleGluIysIysAspGlyThrIleIleProThrLeuValGlnAlaIleGlyGlnAsp 280  
 Db 912 ATATATACAAAGAAAGACCCATCATCTTACACTGTGGAGGCAAGTAAAGAAACAAGAT 971  
 QY 281 GlnArgGluValAspTrpGluTYrPheTYrGlyLeuLeuPheThrSerGluAsnLeuIys 300  
 Db 972 GGAAGAAAGAGTGGAGTATTTTATGGCTGCTTTTACTCAGAGAACTTAA 1031  
 QY 301 LeuValHisIleValCysHisIysLYsLYsThrHisIysLeuAsnCysAspProSerArg 320  
 Db 1032 CTAGTGCACATTTGCTCCATTAAGAAACACCCACAAGCTCAACTGAGACCCAGCAGA 1091  
 QY 321 IleGlyTrpProGlnThrArgLeuArgGlyGlnArgValArgLYsArgGln 338  
 Db 1092 ATCTACAAACCCAGACAGAGTGAAGCGGAAGAGCGCTGTGCGAAACGCCAG 1145

; Sequence 11, Application US/10242943  
 ; Publication No. US20030087412A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Nandabalan, Krishnan  
 ; APPLICANT: Schulz, Vincent P.  
 ; APPLICANT: Yang, Melja  
 ; TITLE OF INVENTION: NIK1 PROTEIN AND NIK1 PROTEIN COMPLEXES  
 ; FILE REFERENCE: 15966-521 NIK1 protein complexes  
 ; CURRENT APPLICATION NUMBER: US/10/242,943  
 ; CURRENT FILING DATE: 2002-09-13  
 ; PRIOR APPLICATION NUMBER: US/09/167,206  
 ; PRIOR FILING DATE: 1998-10-06  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 11  
 ; LENGTH: 1766  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (44)..(1441)  
 ; US-10-242-943-11

## Alignment Scores:

Pred. No.: 0.00537 Length: 1766  
 Score: 102.50 Matches: 77  
 Percent Similarity: 34.42% Conservative: 50  
 Best Local Similarity: 20.87% Mismatches: 128  
 Query Match: 5.73% Indels: 115  
 DB: 9 Gaps: 16

US-09-748-451-2 (1-338) x US-10-242-943-11 (1-1766)

QY 4 LysProLYsSerValIysLeuArgAlaLeuArgSerProArgLYsPheGlyValAlaGly 23  
 Db 546 AAGCCCGCGTGTG---AGTGGAGCGCGCAACCTGGCGGAGAGATATGCGCTCCGGG 602  
 QY 24 ArgSerCysGlnGluValLeuArgLYsGlyCysLeuArgPheGlnLeuProGluArgGly 43  
 Db 603 AGAAATTCAGAG----- 614  
 QY 44 SerArgLeuGlyLeuTYrGluAspGlyTYrGluLeuThrGluLeuArgTYrPheProSerVal 63  
 Db 615 ---AGGAGATCTTCAGAGAGAG--GAAGCCGAAACACCTCGCAATCTTTCAGACAGAT 670  
 QY 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlnGlnAlaTrpGlnGlyTYrValSer 83  
 Db 671 GTTACATATGCGTCTGTGCGACGCTTGACCTTGACGCAAGAGAAATCTTTCAGAGAA 730  
 QY 84 AspleuArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaIa 103  
 Db 731 GAGATT---GCCTTTTCAGAGAACTCCACAGAGAGAAATCCAGAGCTGACGCTCAG 787  
 QY 104 -----GlnGlnLeuLeuCysAspGlnGlnAlaProGlnArgGlnArgLeu 118  
 Db 788 ATTACAGAACAGCATGTCCAAATGATGATGTGATGTTTCCAAAGCTGACCTACGCGTCC 847  
 QY 119 LeuAlaAspleuLeuHisAsnValSerGlnAsnIleAlaIaGluThrArgAlaGluAsp 138  
 Db 848 CTGGCTGCGTAGCTCAGCAATAT---GAAGTGTGGCTGCCAAGAACCTGCAGAGGACA 904  
 QY 139 ProProTrpPhe-----Glu 143  
 Db 905 GAAGAAATGTCATCAATCCAACTTGTGACCTCTGTGAGGCTGCCAACCGGACAAATGAC 964  
 QY 144 GlnLeuGlnSerArgPheGlnSerIysSerGlyTYrLeuArgTYr----- 158  
 Db 965 GCCCTGGCGCCAGGCAAGCAGAGAGTCCACTAGTACCGGAGACAGGTGACCTCCCTCAC 1024  
 QY 159 -----SerCysGlnSerArgIleArg----- 165  
 Db 1025 TGTGAGTGTATGCCCTTAAAGGAACCAATGATGCTCGAAGAGCCAGATGCTGAATATG 1084



```

: RESULT 4
: US-09-974-298-43
: Sequence 43, Application US/09974298
: Patent No. US20020156263A1
: GENERAL INFORMATION:
: APPLICANT: Chen, Hui-Mei
: TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
: FILE REFERENCE: PA-0037 P
: CURRENT APPLICATION NUMBER: US/09/974,298
: CURRENT FILING DATE: 2001-10-04
: PRIOR APPLICATION NUMBER: 60/238,331
: PRIOR FILING DATE: 2000-05-10
: NUMBER OF SEQ ID NOS: 194
: SOFTWARE: PERL Program
: SEQ ID NO 43
: LENGTH: 1864
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incyte ID NO. US20020156263A1 127987.19
: 9-974-298-43

Alignment Scores:
Pred. No.: 0.00584 Length: 1864
Score: 102.50 Matches: 77
Percent Similarity: 34.42% Conservative: 50
Best Local Similarity: 20.87% Mismatches: 128
Query Match: 5.73% Indels: 115
DB: Gaps: 16

US-09-748-451-2 (1-338) x US-09-974-298-43 (1-1864)

OY 4 LysProlySservAllysLeuArgAlaLeuArgSerProAlGylsPheGlyValAlaGly 23
    ||||| ||| ||| ||||| ||||| |||||
Db 642 AAGCCCGGCTG--AGGTGGAGCGCCGCAACCTGGCCGAGACATCATCGCCCTCCGG 698

OY 24 ArgSerCysGInglValLeuArgGlyCysLeuArgPheGInLeuProGluArgGly 43
    ||||| ||||| ||||| ||||| ||||| |||||
Db 699 AGAAATTCAG----- 710

OY 44 SerArgLeuCysLeuTyGluAspGlyThrGluLeuThrGluAspTyR-PheProSerVal 63
    ||| ||| ||| ||| ||| ||| |||
Db 711 --AGGAGATGCTTCACAGAGAG--GAAGCCGAAACACCCTGCATCTTTCAGACAGAT 766

OY 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaIatPpGInGlyTyValSer 83
    ||||| ||||| ||| ||| ||| |||
Db 767 GTTGACATGAGCGTCTGGACGCTGTGACCTTGAAACCAAGTAGGAATCTTGGCAGAA 826

OY 84 AspIleArgArgPheLeuSerAlaPheHisGluProGInValGlyLeuIleGlnAlaIa 103
    ||||| ||||| ||||| ||||| ||||| |||||
Db 827 GAGAT---GCCCTTTTGAAGAAATCCACAGAGAAAGAAATCCAGAGCTCAGGCTCAG 883

OY 104 -----GInGInLeuLeuCysAspGInGlnAlaProGInArgGlnArgLeu 118
    ||||| ||||| ||| ||| |||
Db 884 ATTCAGAAACGATGTCCTCAATGATGATGATGTTTCCAGACCTGCACCTCAGCGCTCC 943

OY 119 LeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrArgAlaGluAsp 138
    ||| ||||| ||||| ||||| ||||| |||||
Db 944 CTGGCTGACGTACGTACGACCAATAT--GAAAGTGTGCTGCCAAGAACCTCAGAGAGCA 1000

OY 139 ProProTrpPhe-----Glu 143
    |||||
Db 1001 GAAGATGCTCAATCCAAATCTTGTACCTCTCTGAGGCTGCCAACCGAACAATATAC 1060

OY 144 GlyLeuLusArgPheGInSerLysSerGlyTyLeuArgTyR----- 158
    ||| ||| ||| ||| |||
Db 1061 GCCCTGGCCAGCAAGACAGAGAGTCACTGATACCGAGACAGAGTCACTCCCTCAC 1120

OY 159 -----SerCysGInSerArgIleArg----- 165
    ||| ||| ||| ||| |||
Db 1121 TGTCAGTGTAGCCCTTAAAGAACCAATGATGATGTCCTGGAAGCCCAATGCTGGAATG 1180

```

Qy	166	-----SetTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGln183
Db	1181	GAAGAGAACTTGGCCGCTGAAAGCTGCTAATCACAAGACATATGGC---CGCCTGCAG1237
Qy	184	GluGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrAsn203
Db	1238	GATGAGATTACAGAAATATGAAGAGAGAAATGGCTGCACCTTGCTGAATACCA-----1291
Qy	204	GlySerTyrPheAspArgGlyAlaIleGlyGlySerArgLeuGlyCysThrProGluGlyTyr223
Db	1292	-----GACCTG1297
Qy	224	PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsnPro243
Db	1298	CTCAATGTTAAGATGGGCCCTTGACATTTGAGATTGCCACCTACAGAAAGCTGCTG-----1351
Qy	244	TyrSerAsnArgGluSerArgIleLeu-----PheSerThr-----255
Db	1352	---GAAGCGGAGAGAGAGAAATTTCTGCTCCTTCCAAACTTTTCTCCTCCAGACCTG1408
Qy	256	-----TyrAsnLeuAspHisIle-----IleGluIleGlySerArgThrIleIle269
Db	1409	ACGGAAACTAATCTGGATTCTACCTCCCTGTTGATACCACTCAAAAGAGACTT---1465
Qy	270	ProThrLeuValGluAlaIleIleGlyGluGlnAspGlyArgGluValAsp-----285
Db	1466	-----CTATTAAGACAGCGTTGGAACAGAGATGGACAGGTATCAACAGAACTTCTCAG1519
Qy	286	-----TyrGluTyrPheTyrGlyLeuLeuPhe-ThrSerG1297
Db	1520	CATCAGATGACCTTGAAATTAATAATTGCACACACTCAGTCAGACAAATATATTACCAACAA1579
Qy	297	uAsnLeuIleValHisIleVal1305
Db	1580	GAATTAATAAAGAAATCCATATCTTTA1604
RESULT 5		
US-10-044-090-793		
Sequence 793, Application US/10044090		
Patent No. US20020137081A1		
GENERAL INFORMATION:		
APPLICANT: Olga Bandman		
TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION		
FILE REFERENCE: PA-0028 US		
CURRENT APPLICATION NUMBER: US/10/044, 090		
CURRENT FILING DATE: 2002-01-09		
NUMBER OF SEQ ID NOS: 850		
SOFTWARE: PERL Program		
SEQ ID NO 793		
LENGTH: 1892		
TYPE: DNA		
ORGANISM: Homo sapiens		
FEATURE:		
NAME/KEY: misc_feature		
OTHER INFORMATION: Incyte ID No. US20020137081A1 2813002CBI		
US-10-044-090-793		
Alignment Scores:		
Pred. No.:	0.00598	Length: 1892
Score:	102.50	Matches: 77
Percent Similarity:	34.42%	Conservative: 50
Best Local Similarity:	20.87%	Mismatches: 128
Query Match:	5.73%	Indels: 115
DB:	12	Gaps: 16
US-09-748-451-2 (1-338) x US-10-044-090-793 (1-1892)		
Qy	4	IysProIysSerValIysLeuArgAlaLeuArgSerProArgIysPheGlyValAlaGly23
Db	657	AAGCCCGGCGTGC---AGTGGAGAGCGGAGACAACTGGCCGAGAGACATATCGCTCCGGG713
Qy	24	ArgSerCysGlnGluValLeuArgIysGlyCysLeuArgPheGlnLeuProGluArgGly43

D	b		714 AGAATTTCAGAG-----	725
O	y	44	SerATgLeuCysLeuATgTgtGluAspGlyThrGluLeuThrGluAspTyrrPheProSerVal	63
D	b	726	---AGAGAGTGCTTCAGAGAG- GAAGCGCAAAACACCCTTCATCTTTCCAGACAGAT	781
O	y	64	ProARAsnAlaGluLeuValLeuLeuThrIleuGlyGlnAlaLysTrpGlnGlyTyrValSer	83
D	b	782	GTTGCATATGGCTGTCTGCCAGCCTTGACCTTGAACGAAAGTAGAATCTTTCACAAGA	841
O	y	84	AspILeATgATgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaLa	103
D	b	842	GAGATr---GCCCTTTTGAAAATCTCCAGAAAGGAAATCCAGACTCCAGGCTCAG	898
		104	-----GlnGlnLeuLeuCysAspGluGlnAlaLysProGlnAlaTrpGlnAla	118
		899	ATTTCAGAACACAGCATGTCTCAAATCGATGTGGATGTTTCCAAACCTTCAGCTCCAGCTCC	958
O	y	119	LeuAlaAspLeuLeuHisAsnValLysGlnAsnIleAlaAlaGluThrArgAlaLysP	138
D	b	959	CTGCGTGACGTAAGTCAGCAATAT--GAAGATGTGGCTGCCAAGAACCTTCAGAGAGCA	1015
O	y	139	ProPTOTPrHe-----	143
D	b	1016	GAAGAATGATCAATAATCCAAGTTTGCTGACCTGTGAGGCTGCCAAGCGGAACAATGAC	1075
O	y	144	GlyLeuGluSerATgArpPheGlnSerLysSerGlyTyrLeuArgTyr-----	158
D	b	1076	GCCCTGCCGCGCAAGACAGAGATCCACTGACTACCGGAGACAGTGCACCTCCCTCAC	1135
O	y	159	-----SerCysGluSerATgIleArg-----	165
D	b	1136	TGTGAATGTGATGCCCTTAAAGAACCAATGATGCCCTGGAAAGCGCAATGCGTAATG	1195
O	y	166	-----SerTyrLeuATgGluValLysSerTyrProSerTyrValGlyAlaGln	183
D	b	1196	GAAGAACTTGGCCGTGAAGAGCGCTACCTACCAAGACACTATTGGC---CGCTGCAG	1252
O	y	184	GluGluPheLeuATgValLeuGlySerMetCysGlnAlaTrpLysSerMetGlnTyrAsn	203
D	b	1253	GATGAGATTCAGATATGAAGAGAAATGCTGTGCACCTGTGGATACCA-----	1306
O	y	204	GlySerTyrPheAspATgGlyAlaLysGlyLysSerATgLeuCysThrProGluGlyTyr	223
D	b	1307	-----GACCTG	1312
O	y	224	PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerATgHisSerIleAsnPro	243
D	b	1313	CCTCATGTGTAAGATGGCCCTTGAACATTGACATGGCCACCTACAGAACCTGTG---	1366
O	y	244	TyrSerAsnATgGluSerATgIleLeu-----PheSerThr-----	255
D	b	1367	---GAAGCGAGAGAGAGAGATTTCTCTCTCTCCAAATTTTTTCTCCCTCCAGACCTG	1423
O	y	256	-----TrpAsnLeuAspHisIle-----IleGluLysLeuArgThrIle	269
D	b	1424	AAGGAATATATCTGATTCACCTCCCTGTGATGATCCCAAAAAAGACACT---	1480
O	y	270	ProTHrLeuValGlnAlaLysGluGlnAspGlyATgGluValAsp-----	285
D	b	1481	-----CTGATTAAGACCGTTGAACACTAGAGATGAGACAGCTTATCAAGAACTTTCAG	1534
O	y	286	-----TrpGluTyrPheTyrGlyLeuLeuPhe-ThrSerGI	297
D	b	1535	CATCAAGATGACCTTGAATAAAAATTCGACACACTAGTCAGCAATATATTCACAGCA	1594
O	y	297	uasnLeuLysLeuValHisIleVal	305
D	b	1595	GAATAAAAAGAAATCCATATCTTA	1619

Sequence 44, Application US/09974298  
Patent No. US20020156263A1  
GENERAL INFORMATION:  
APPLICANT: Chen, Hwei-Mei  
TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER  
FILE REFERENCE: PA-0037 P  
CURRENT APPLICATION NUMBER: US/09/974,298  
PRIOR FILING DATE: 2001-10-04  
PRIORITY DATE: 2000-05-10  
NUMBER OF SEQ ID NOS: 194  
SOFTWARE: PERL Program  
SEQ ID NO 44  
LENGTH: 3405  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/key: misc.feature  
OTHER INFORMATION: Incyte ID No. US20020156263A1 1397781.7  
US-09-974-298-44

Alignment Scores:

Pred. No.:	0.015	Length:	3405
Score:	102.50	Matches:	77
Percent Similarity:	34.42%	Conservative:	50
Best Local Similarity:	20.87%	Mismatches:	128
Query Match:	5.73%	Indels:	115
DB:	9	Gaps:	16

US-09-748-451-2 (1-338) x US-09-974-298-44 (1-3405)

O	y	4	LysProLysSerValLeuATgValLeuATgSerProATgLysPheGlyValAlaGly	23
D	b	784	AAACCCGCTCG---AGGTGAGCGCGCAACCACTGGCCAGAGACATCATGGCTCCGG	840
O	y	24	ArgSerCysGlnGlnValLeuATgValGlyCysLeuATgPheGlnLeuProGlnATgLy	43
D	b	841	AGAAATTCAGG-----	852
O	y	44	SerATgLeuCysLeuTgtGluAspGlyThrGluLeuThrGluAspTyrrPheProSerVal	63
D	b	853	---AGGAGATGTCTCAGAGAGAG-GAAGCGAAACACCCGCAATCTTTGACAGAGAT	908
O	y	64	ProARAsnAlaGluLeuValLeuLeuThrIleuGlyGlnAlaLysTrpGlnGlyTyrValSer	83
D	b	909	GTTGACAAATGCTGTCTGTGCACCTGTGACCTTGAACGCAAGATGGAATCTTTGCAAGAA	968
O	y	84	AspILeATgATgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaLa	103
D	b	969	GAGATr---GCCCTTTTGAAAATCTCCAGAAAGGAAATCCAGAGACTCCAGGCTCAG	102
O	y	104	-----GlnGlnLeuLeuCysAspGluGlnAlaLysProGlnAlaTrpGlnAla	118
D	b	1026	ATTTCAGAACACAGATGTCTCAAATCGATGTGGATGTTTCCAAACCTTCAGCTCCAGCTCC	108
O	y	119	LeuAlaAspLeuLeuHisAsnValLysGlnAsnIleAlaAlaGluThrArgAlaLysP	138
D	b	1086	CTGCGTGACGTAAGTCAGCAATAT--GAAGATGTGGCTGCCAAGAACCTTCAGAGAGCA	1144
O	y	139	ProPTOTPrHe-----	143
D	b	1143	GAAGAATGTGAACATTCGAATTTGCTGACCTGTGAGGCTGCCAAGCGGAACAATGAC	120
O	y	144	GlyLeuGluSerATgArpPheGlnSerLysSerGlyTyrLeuArgTyr-----	158
D	b	1203	GCCCTGCCGCGCAAGACAGAGATCCACTGACTACCGGAGACAGTGCACCTCCCTCAC	126
O	y	159	-----SerCysGluSerATgIleArg-----	165
D	b	1263	TGTGAATGATGCCCTTAAAGAACCAATGATGCCCTGGAAAGCGCAATGCGTAATG	1322
O	y	166	-----SerTyrLeuATgGluValSerSerTyrProSerTyrValGlyAlaGln	183



```

GENERAL INFORMATION:
APPLICANT: HYBRIDGENICS
TITLE OF INVENTION: Protein-protein interactions between Shigella flexneri polypeptides
FILE REFERENCE: B4778A
CURRENT APPLICATION NUMBER: US/10/043,487
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/261,130
PRIORITY FILING DATE: 2001-01-12
NUMBER OF SEQ ID NOS: 561
SOFTWARE: Patent version 3.1
SEQ ID NO: 101
LENGTH: 1218
TYPE: DNA
ORGANISM: Shigella flexneri
US-09-748-451-2 (1-338) x US-10-043-487-101 (1-1218)

Alignment Scores:
Pred. No.: 0.00741 Length: 1218
Score: 99.50 Matches: 71
Percent Similarity: 34.23% Conservative: 44
Best Local Similarity: 21.13% Mismatches: 120
Query Match: 5.56% Gaps: 102
DB: 9 Indels: 13

US-09-748-451-2 (1-338) x US-10-043-487-101 (1-1218)
OY 4 LysProLysSerValLysLeuAlaLeuAlaSerProArgLysPheGlyValAlaGly 23
DB 320 AACCCCGCCCTGC---AGGTGAGCGCGCAACCTGGCCGACGACATCATGCGCCGCG 376
OY 24 ArgSerCySGIngluValLeuAlaGlyGlyCysLeuAlaArgPheGlnLeuProGluAlaGly 43
DB 377 AGAAATGTCAGG----- 388
OY 44 SerArgLeuCyLeuTyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerVal 63
DB 389 ---AGGAGATGCTTCAGAGAG- GAAGCGAAGAACACCTGCATCTTCACAGAGAT 444
OY 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTyrPheGlnGlyValSer 83
DB 445 GTTGACAAATGCGCTCTGCGACGCTTGACCTTGACGACGAAAGTGGATCTTTCAGAA 504
OY 84 AspLeuAlaGlyArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAla 103
DB 505 GAGATT---GCTTTTAAAGAAACCTCCAGAAAGAGAAATCCAGAGAGCTGAGGCTCAG 561
OY 104 -----GlnGlnLeuLeuCyAspGluGlnAlaProGlnArgGlnArgLeu 118
DB 562 ATTCAGAAACAGCATGTCCAAATCATGATGTGATGTTCCAAAGCTGACCTCAGAGCTGCC 621
OY 119 LeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaAlaGluThrArgAlaGluAsp 138
DB 622 CTGCGTGCAGCTGACGTCAGCAATAT- GAAGTGTGCTGCTGCCAAGAACCTGCGAGAGGCA 678
OY 139 ProProTyrPhe-----Glu 143
DB 679 GAAGAATGGTCAAAATCCAAATGTTGACCTCTGAGGCTGCCAACCGGACAAATGAC 738
OY 144 GlyLeuGluSerArgPheGlnSerLysSerGlyTyrLeuAlaTyr----- 158
DB 739 GCCCTGCGCAGCAAGACAGATTCATGATACGAGGACAGAGTGCAGCTCAGC 798
OY 159 -----SerCySGIngluValLeuAlaGly----- 165
DB 799 TGTGAGTGTGATCCCTTAAAGAACCAATGATCCCTGCAACGCCCAATGCTGTAATG 858
OY 166 -----SerTyrLeuAlaGluValSerSerTyrProSerThrValAlaGluAlaGln 183
DB 859 GAAGAGAACTTTCGCTGTAAGCTGCTACACCAAGACACATATGTC---GCGCTGAG 915
OY 184 GluGluPheLeuAlaValLeuGlySerMetCySGIngluArgLeuArgSerMetLeuTyrAsn 203

```

```

DB 916 GATGAGATTCACATATATAGAGAGAAATGCTGCTGACCTTGGAATACCA----- 969
OY 204 GlySerTyrPheAspArgGlyAlaLysGlySerArgLeuCyThrProGluGlyTyr 223
DB 970 -----GACCTG 975
OY 224 PheSerCySGIngluTyrPheAspMetAspSerCySGIngluSerArgHisIleAsnPro 243
DB 976 CTCATATGTAATGATGCGCTTGACATGATGATGACCTACAGAACCTGCTG----- 1029
OY 244 TyrSerAsnArgLysSerArgIleLeu-----PheSerThr----- 255
DB 1030 ---GAAGCGAGAGAGAGAGATTTCTGCTCTTCCAAATCTTCTCCCTGCAACCTG 1086
OY 256 -----TyrAsnLeuAspHisIle-----IleGluLysLysArgThrIleIle 269
DB 1087 AGGGAACATATCTGATTCACCTCCCTGCTGATGATCCCACTCAAAAAGACATTC--- 1143
OY 270 ProThrLeuValGluAlaIleLysGluGlnAspGlyArgGluValAsp 285
DB 1144 -----CTCATTAACGCGTTGAACATGAGATGACAGATTATAC 1185

RESULT 9
US-10-161-803-30
Sequence 30, Application US/10161803
Publication No. US20030092028A1
GENERAL INFORMATION:
APPLICANT: Ma, Yuanhong
APPLICANT: Li, Chih-jian
APPLICANT: Chen, Fan
APPLICANT: Palman, Jeffrey
TITLE OF INVENTION: TREATMENT OF INSULIN RESISTANCE AND RELATED CONDITIONS
FILE REFERENCE: 42145200300
CURRENT APPLICATION NUMBER: US/10/161,803
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/295,264
PRIOR FILING DATE: 2001-06-01
NUMBER OF SEQ ID NOS: 61
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 30
LENGTH: 2209
TYPE: DNA
ORGANISM: Homo sapiens
US-10-161-803-30

Alignment Scores:
Pred. No.: 0.0188 Length: 2209
Score: 99.50 Matches: 77
Percent Similarity: 33.14% Conservative: 39
Best Local Similarity: 22.00% Mismatches: 127
Query Match: 5.56% Indels: 108
DB: 9 Gaps: 16

US-09-748-451-2 (1-338) x US-10-161-803-30 (1-2209)
OY 4 LysProLysSerValLysLeuAlaLeuAlaSerProArgLysPheGlyValAlaGly 23
DB 585 AACCCCGCCCTGC---AGGTGAGCGCGCAACCTGGCCGACGACATCATGCGCCGCG 641
OY 24 ArgSerCySGIngluValLeuAlaGlyGlyCysLeuAlaArgPheGlnLeuProGluAlaGly 43
DB 642 AGAAATGTCAGG----- 653
OY 44 SerArgLeuCyLeuTyrGluAspGlyThrGluLeuThrGluAspTyrPheProSerVal 63
DB 654 ---AGGAGATGCTTCAGAGAG- GAAGCGAAGAACACCTGCATCTTCAGACAGAT 709
OY 64 ProAspAsnAlaGluLeuValLeuLeuThrLeuGlyGlnAlaTyrPheGlnGlyValSer 83
DB 710 GTTGACAAATGCGCTCTGCGACGCTTGACCTTGACGCAAGAGTGAATCTTTCAGAA 769

```

```

QY      84  AsplLeaArgArpHeLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaAla 103
      770  GAGATT---GCCCTTTTGAAGAAACTCCAGAGAGAGAAATCCAGAGAGCTGCAGCTCAG 826
QY      104  -----GlnGlnLeuLeuCyAspGlnGlnAlaPheGlnAlaPheGlnArgLeu 118
      827  ATTCAGACACAGCATGTCCTCCAAATCGATGTGATGTTTCCAAAGCTCAGCTCAGCGCTGCC 886
QY      119  LeuAlaAspLeuLeuHisAsnValSerGlnAsnIleAlaIleGlnThrArgAlaGluAsp 138
      887  CTGGGTGACGTACGTACGACATAT---GAAAGTGTGCTGCCAAGAACCTGCAGAGAGCA 943
QY      139  ProGlnPhe-----Glu 143
      944  GAAGAAATGTACAAATCCAAATGTTGCTGACCTCTGTGAGGCTGCCAACCGAACATATAC 1003
QY      144  GlyLeuGlnSerArgPheGlnSerGlyTyrrLeuArgTyrr----- 158
      1004  GCCCTGCCGACGACCAAGCAGAGCTCCAGTACGAGACAGCTGCAGCTCCTCACC 1063
      159  -----SerCysGlnSerArgGlyLeu----- 165
      1064  TGTGAATGTGATGCCCTTAAGAACCAATGATGATGCTGAGAGCCAGATGCGTGAATG 1123
QY      166  -----SerTyrrLeuArgGlnValSerSerTyrrProSerThrValGlyAlaGln 183
      1124  GAAGAGAACTTTCCTGCTGAAGCTGTACTACTACCAAGACATATTTGCC---CGCCTCAG 1180
QY      184  GlnGlnPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrrAsn 203
      1181  GATGAGATTTCAGATATGAAAGAGAAATGCGTCGTACCTTCGTGAATACCA--- 1234
QY      204  GlySerTyrrPheAspArgGlyAlaValGlyGlySerArgLeuGlyThrProGlnGlyTyr 223
      1235  -----GACCTG 1240
QY      224  PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsnPro 243
      1241  CTCAAATGTTAAGATGCGCTTGACATGATGATGACACCTACAGAGAACTGCTG--- 1294
QY      244  TyrSerAsnArgGlnSerArgIleLeu-----PheSerThrTyrPasnLeu 258
      1295  ---GAAGCGACGAGAGACAGATTTCTGCTCCTTCCAAACTTTTCTCCGCAACCTG 1351
QY      259  -----AspHisIleIleGlu-----LysAlaArgThrIleIleProThrLeuValGlu 274
      1352  AGGCGTAAAGCATTTTATTTCCCTTTAGGAAAAAGCTCAGCTGCTGTAACACAGCTGTTT 1411
QY      275  AlaIleGlyGlnGlnAspGlyArgGlnValAspTyrGlnTyrrPheTyrrGlyLeuLeuPhe 294
      1412  ATGTCAAAAGCAT-----TCATTTTATTTAGATAT----- 1441
      295  ThrSerGlnAsnLeuLeuValHisIle 304
      1442  -----CTGAAAAAATGCCAATATA 1459

```

```

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)---(69327)
; OTHER INFORMATION: n = A,T,C or G
US-09-777-921A-3

Alignment Scores:
Pred. NO.: 4.08 Length: 69327
Score: 99.50 Matches: 86
Percent Similarity: 39.69% Conservative: 66
Best Local Similarity: 22.45% Mismatches: 123
Query Match: 5.56% Indels: 108
Gaps: 23

US-09-748-451-2 (1-338) x US-09-777-921A-3 (1-69327)
QY      2  LeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyVal 21
      36404  ATGCAGAAAGCCTCAGAGCTG-----ATGCATCACTGGAGAAAGGCTATC 36451
QY      22  Ala---GlyArg-SerCysGlnGlnValLeuArgGlyCysLeuArgPheGlnLeuP 40
      36452  AGCGATGAGAGATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 36511
QY      40  OGluArgGlySerArg-----LeuCy 47
      36512  AAAAGAAAGAGAGAAAGCCTCCAAAGAAATATGAGCATATGTGAAAGAACCAATCATATG 36571
QY      47  GLeuTyrr-----GluAspGlyThrGlnLeu----- 55
      36572  TCTGATTTGTTACTGAAAGTGAAGTGAAGGAGAAATGAAACAGTTGAAACACTCTGCA 36631
QY      56  -----ThrGlnAspTyrrPheProSerValProAspAsnAlaGlnLeuValLeuLeuTh 73
      36632  GGATATATTCAGAGAGAACTTCCCAATCTTACAGACAGCAGCAGCAATTCAGATTCAGCA 36691
QY      73  rLeuGlyGlnAlaTrpGlnGlyTyrrValSerAspIleArgArgPheLeuSerAlaPhe 93
      36692  AATACAGAGAAAGCCAGCAAGATATC-----TCCTTGAGAAAGCACTCCAAAGACAT 36745
QY      93  SGlnProGlnValGlyLeuIleGlnAlaAlaGlnGlnLeuLeuCysAspGlnGlnAlaP 113
      36746  AATTGTCAAGATTCACCAAGTGTGAATGAAGAAATGTTAAGGGCAGCAGAGAGAA 36805
QY      113  OGlnArgGln-----ArgLeuAlaAspLeuHisAsnVal 126
      36806  AGTGGGTTACCTCAATGGAAGCCATCAGACTACAGCGGATCTTTGGCAGAAAC 36865
QY      126  lSerGlnAsnIleAlaAlaGlnThrArgAlaGluAspProTyrPheGlnGlyLeu-- 145
      36866  TCTACAA-----ACCAAGAAAGAGTGGGGCAATATTCATTTCTTA 36910
QY      146  -GlnSerArgPheGlnSerLysSerGlyTyrr-----LeuArgTyrrSerCysGln 162
      36911  AGAAAGAAATTTTCACCCAGAAATTCATATCCAGCCAACTAAGCTTCAATGAAGAG 36970
QY      162  rArgIleArgSerTyrr-----LeuArgGlu---ValSerSerTyrrProse 176
      36971  AGAAATTAATTCCTTTCACAGCAAGCAAAATGCTGAGAGATTTTGTCCACCGAGCTT-- 37028
QY      176  rThrValGlyAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 196
      37029  -----GCCCTAAAGAGTCTCGAAG-----GA 37051
QY      196  GLeuArgSerMetGlnTyrrAsnGlySerTyrrPheAspArgGlyAlaLysGlyLysSer 216
      37052  AGTCTTAATCTGGAAGAAACATACATGATC-----CAAGCCCTGCAAAATATGCGCA 37105
QY      216  GLeuCyThrProGlnGlyTyrrPheSerCysGlnGlyProPheAspMetAspSerCys 236
      37106  AATGTAAAGACG-----TCGACACT 37126

```

Db	979	GAGAGT---	GCCTTTTGGAAAGACTCCAGCAAGACAGAAATCCAGAGACTGCAAGCTCAG	1035
Oy	104	-----	glnlnleuleucysarglunlnalprogllnaraglnargleu	118
Db	1036	ATTGACGACACAGCATGTCTCCAAATGCGATGTGGATGGTTTCCAAAGCTCAGCTCCAGCTGCC	1095	
Oy	119	leuAlAspleuLeuHlshsAnvAlserGlnAsnIleAlaIagIuNrArAlaIagIuAsp	138	
Db	1096	CTGGGTATCTACCTGTCAGCCAAATAT--	GAAGTGTGGTCGCAAGAACTCGCAGGAGCCA	1152
Oy	139	ProTrOTrphe-----	glu	143
Db	1153	GAAAGAAAGTTCACAAATCCAAAGTTTCTGACCTCTCGAGAGCTGCCAACCGGAACAAATGAC	1212	
Oy	144	GlyleuGlnSerArpPheGlnSerIysSerGlyTYrLeuArTYr-----	158	
Db	1213	GCCCTGGCCAGCAAGACAGAGAGTCACACAGTAATACCGGAGACAGGTGACGTCCCTACCC	1272	
Oy	159	-----	SerCysGlnSerArpGlyleu	165
Db	1273	TGTGAATGATGATCCCTTAAAGAACCAATGATCCCTTGAAAGCCGACAGTGGTGAATG	1332	
Oy	166	-----SerTYrleuArGlnAlaIserSerTYrProSerThrValGlyAlaIagln	183	
Db	1333	GAAAGAACTTTCCTCCCTGTAAGCTGCTACTACTACCAAGACACTATTTGG--	CGCCTGCAG	1389
Oy	184	GlnGlnPheLeuArTYrAlleuGlySerMetCysGlnIArgLeuArGSerMetGlnTYrAsn	203	
Db	1390	GATGAGATTTCAGATTTGTGAAGAGAGAAATGGCTGTCACCTTCGTAATACCAA-----	1443	
Oy	204	GlySerTYrPheAspArgIyAlaIagIyGlySerArGleuCYthrProGlnGlyTYr	223	
Db	1444	-----	GACCTG	1449
Oy	224	PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHIsSerIleAsnPro	243	
Db	1450	CTCAATGTAAATGATGGCCCTTGACATTGAGATTGCCACCTACAGAAAGCTGCTG--	1503	
Oy	244	TYrSerAsnArGlnSerArpGlyleu-----	PheSerThr	255
Db	1504	---GAAAGCCAGAGAGACAGATTTCTCTGCTCTTCACAAACTTTCTCTCCGCAACCTG	1566	
Oy	256	-----TrpAsnleuAspHisIle-----	IleGlnIysIysArpThrIleIle	269
Db	1561	AGGAAACTAATCTGGATCTACCTCCCTGCTGGTATACCCACTCAAAAGACATTC--	1617	
Oy	270	ProThrleuValGlnAlaIleIysGlnGlnAspGlyArGlnAlaAsp	285	
Db	1618	-----CTGATTAAGACGGTTGAAACTAGAGATGACAGCTTATCAAC	1659	
RESULT 12				
US-10-060-425-7/c				
Sequence 7, Application US/10060425				
Patent No. US2002016450A1				
GENERAL INFORMATION:				
APPLICANT: Hiebsch, Ronald				
TITLE OF INVENTION: Methods of Assessing Wolfram Protein Activity				
CURRENT REFERENCE: 00450.US1				
CURRENT APPLICATION NUMBER: US/10/060,425				
CURRENT FILING DATE: 2002-01-30				
PRIOR APPLICATION NUMBER: 60/266,385				
PRIOR FILING DATE: 2001-02-02				
NUMBER OF SEQ ID NOS: 17				
SOFTWARE: PatentIn version 3.1				
SEQ ID NO: 7				
LENGTH: 2673				
TYPE: DNA				
ORGANISM: Rattus norvegicus				
US-10-060-425-7				
Alignment Scores:		0.0623		Length: 2673



Oy		30	Leu-----	-ArgLysGlyCysLeuAcgPrphglnIleuProGluYr	42
Dd		430	GTAACAGGCACCGCAATGGCAAGCATAGGTCCAACTTCACTGCAACTTCACATTTCACAACTTCCAAACCA	489	
Oy		43	GlySerArgLeucys-----	-LeuTrgluAspGlyThrTrgluLeu	55
Dd		490	GGCATTCGAATTTTGCTGTTGCTGTTCTGTGTCAACAACCTGTGTGAACCTGTGTCAACGGAGTGT	549	
Oy		56	ThrgluAspTrpPheProSerValProAspAlaGluLeu-----	-Val	70
Dd		550	CGTCCGGCGCTCATCTCCCGCTGCCAAGCCGCAACCGCTTTTCAGAACAAGTTCCTCCCGG	609	
Oy		71	LeuleuthrLeuGlyGlnAlaTrpGlnGlyTyrrValserAspIleArgArgPheLeuSer	90	
Dd		610	GTGTGTGAACCGGGCTGCGAACCGCGAGCGCTTT-----	-CGTTTTGNG--	651
Oy		91	AlArphenlaGluProGlnValGlyLeuIleGlnAlaIaGlnIleuLeuCysAspGlu	110	
Dd		652	GTTTTTTCTGGGAACAGATGGSGATTCCTTCGATGGCGAAGCGTTG	699	
Oy		111	GlnAlaProGlnArgGlnArgLeuAlaAspLeuLeuHisAsnValSerGlnAsnIle	130	
Dd		699	-----	699	
Oy		131	AlAlaGluThrArgAlaGluAspProTrpPheGlnGlyLeuGluSerArgPheGln	150	
Dd		700	CCTCTGTTCACAGCGCGGTGAAGTCCG-----	-GAACCGTTCACCTTCAGG--	744
Oy		151	SerLysSerGlyTyrrLeuArg-----	-TyrSerCysGluSerArgIleAcgSerTyrrLeu	168
Dd		745	GCGAGGAGCGAGATACCTGCTGTGATATACCTTCGTCGGGCGGGTGCCTCA-----	798	
Oy		169	ArgGluValSerSerTyrrProSerThrValGlyAlaGluAlaGlnGluGluPheLeuArg	188	
Dd		799	-----	-GCGGGTTTTCTGCGAGGAGATCTCGAAG	828
Oy		189	ValLeuGlySerMetCysGlnArg-----	-LeuArgSerMetGlnTyrr	202
Dd		829	ACTTCAGGCTCACAGTGTACACAGATCTGATGTGCATGGCCATTTGCGGCATTC	882	
Oy		203	AsnGlySerTyrrPhasprArgLy-----	-AlaLysGlyGlySerArgLeu	217
Dd		883	-----	-TCTGTTTCCGTCGTGTGCGCANGTTCACACAGACCCTGTGATGAGACCGCC	936
Oy		218	CysThrProGluGlyTyrrPheSerCysGlnGlyProPheAspMetAspSerCysLeuSer	237	
Dd		937	ACGAGCCCCGCAAGCGGTGGCTG-----	-CCTCTT	963
Oy		239	ArgHisSerIle-AsnProTyrrSerAsn-----	-ArgGl	248
Dd		964	TCAACGCGACATCAAAACCCGAGAACAACTCTCAGAAAGCTCGCGGAAAAACGAGGGA	1023	
Oy		248	UserArgGlleuPheSerThrTrpPheLeuAspRhisIleIleGlyLysArgThrI1	268	
Dd		1024	ACAAGAGCGCGTAGAAGTGCACGAAAACCTGAGTGTGTGCGTATGTAGACAAAGTGAAT	1083	
Oy		268	eIleProthrLeuValGlnAlaIleLysGlnGlnAspArgTyrrGluValAspTrpLeuYr	288	
Dd		1084	CTTCCCAAGATGTGCT-----	-CGGATGTGTGCGTGATGCT	1119
Oy		288	rPheTyrglyLeuLeuPheThrSerGluAsnLeuLysLeuValHisIleValCysHisLy	308	
Dd		1120	GTCAACGTGAGATCTGTGCTTCGTCAAAGACCTGTGCTGACCCATCTGCTCTCCATAAA	1179	
Oy		308	slyStrThrHisLysLeuAsnCysAspProSerArgIleTyrrLysProGlnThrArgLe	328	
Dd		1180	AAAGGTT-----	-GTTAACGGCTCTATAAAAAAAAAA	1209
Oy		328	ULysArgLysGln	332	
Dd		1210	AAAAAAAAAAAAA	1222	









Search completed: May 26, 2003, 16:58:10  
Job time : 230 secs

```
OY 154 GlyTyrIleuArgTyrSerCysGluSerArgIleuArgSerTyrIleuArgGluValSerSer 173
Db 6102 ACTCGATTTGTTTCATGCGAAGCAGGAGTGTGCTGACACAGGCCAGAGTGAGCTCC 6043
OY 174 TyrProSerThrValGluAlaGluAlaGluInlu---GluPheIleuArgValIleuGlySer 192
Db 6042 ---CCATCTACTGCTGGACGACAGGCGACCTTCTCTCCAGGCTTCCACCCACTGGGGAGC 5986
OY 193 MetCysGluArgLeuArgSerMetGlnTyrAsnGlySerTyrPheAspArgGlyAlaLys 212
Db 5985 CTTTGT-----ACATTCGAAAGTGAACCAAGTCATGGATGAA-----CAG 5944
OY 213 GlyGlySerArgLeuCysThrProGluGlyTyrP 223
Db 5943 GGTGGCCCGAGATTAGCATTCACAGAGGCTTG 5911

RESULT 20
US-09-764-891-9340/C
Sequence 9340, Application US/09764891
Publication No. US20030077808A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC006
CURRENT APPLICATION NUMBER: US/09/764,891
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 10231
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 9340
LENGTH: 11598
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-891-9340

Alignment Scores:
Pred. No.: 4.35 Length: 11598
Score: 90.00 Matches: 36
Percent Similarity: 41.22% Conservative: 18
Best Local Similarity: 27.48% Mismatches: 63
Query Match: 5.03% Indels: 14
DB: 9 Gaps: 5

US-09-748-451-2 (1-338) x US-09-764-891-9340 (1-11598)
OY 94 GluProGlnValGlyLeuIleGlnAlaGlnIleuLeuCysAspGluGlnAlaPro 113
Db 6264 GAACCACAGCGAGGCCGCGTACACACAGGAGTACTTGTCTGTGCTGCCCTAGCCAG 6205
OY 114 GlnArgGlnArgLeuAlaAspLeuIleAsnValSerGlnAsnIleAlaGlu 133
Db 6204 CAGCTGGAGAACCCCTGGCC-----TGTGGAGACGTAGTCAGGGA 6163
OY 134 ThrArgAlaGluAspProTyrPheGlnGlyLeuGluSerArgPheGlnSerLysSer 153
Db 6162 CCCATGTCACAGACAGAGGCCATTGGAGGAAATGGAAGCGATTAGAGTCCAGTTAT 6103
OY 154 GlyTyrLeuArgTyrSerCysGluSerArgIleuArgSerTyrLeuArgGluValSerSer 173
Db 6102 ACTCGATTTGTTTCATGCGAAGCAGGAGTGTGCTGACACAGGCCAGAGTGAGCTCC 6043
OY 174 TyrProSerThrValGluAlaGluAlaGluInlu---GluPheIleuArgValIleuGlySer 192
Db 6042 ---CCATCTACTGCTGGACGACAGGCGACCTTCTCTCCAGGCTTCCACCCACTGGGGAGC 5986
OY 193 MetCysGluArgLeuArgSerMetGlnTyrAsnGlySerTyrPheAspArgGlyAlaLys 212
Db 5985 CTTTGT-----ACATTCGAAAGTGAACCAAGTCATGGATGAA-----CAG 5944
OY 213 GlyGlySerArgLeuCysThrProGluGlyTyrP 223
Db 5943 GGTGGCCCGAGATTAGCATTCACAGAGGCTTG 5911
```

Tue May 27 09:31:58 2003

us-09-748-451-2.rst

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OK protein - nucleic search, using frame\_plus\_p2n model

Run on: May 26, 2003, 15:23:53 ; Search time 1707 Seconds

(Without alignments)  
3206.839 Million cell updates/sec

Title: US-09-748-451-2

Perfect score: 1789

Sequence: 1 MCRKFSYKLRALRSPKRF.....SRVYKPTRLKRPVRRKQ 338

Indexing table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seg length: 0

Maximum DB seg length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

*No. of hits  
1789*

Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_plus\_p2n.model -DEV=xlh  
-O=/gen2\_1/USPTO.spool/US09748451/unat.21052003.153830.16291/app.query.fasta.1.519  
-DB=EST -OPMT=fastap -SUFFIX=1st -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=Dlosum62 -TRANS=human40.cdi -LIST=45  
-DOCLALIGN=200 -THR.SCORE=Pct -THR.MAX=100 -THR.MIN=0 -ALIGN=20 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09748451.ecgn.1.1.1906.gunat.21052003.153830.16291 -NCPu=6 -ICPU=3  
-NO.MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

EST:  
1: em\_estha:  
2: em\_esthum:  
3: em\_estlin:  
4: em\_estlun:  
5: em\_estrov:  
6: em\_estrpl:  
7: em\_estro:  
8: em\_hic:  
9: gb\_est1:  
10: gb\_est2:  
11: gb\_hic:  
12: gb\_est3:  
13: gb\_est4:  
14: gb\_est5:  
15: em\_estfun:  
16: em\_estom:  
17: gb\_gss:  
18: em\_gss\_hum:  
19: em\_gss\_inv:  
20: em\_gss\_pln:  
21: em\_gss\_vrt:  
22: em\_gss\_fun:  
23: em\_gss\_man:  
24: em\_gss\_mus:  
25: em\_gss\_other:  
26: em\_gss\_pro:  
27: em\_gss\_tod:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	969	54.2	104	14	B0715165	B0715165 AGNCOURT
2	957	53.5	796	13	B1657172	B1657172 603280838
3	954.5	53.4	834	12	B6863734	B6863734 602796717
4	871	48.7	685	13	B1517452	B1517452 603041725
5	774	43.3	830	13	B1770511	B1770511 603055488
6	761	42.5	855	12	BP165515	BP165515 601777413
7	738	41.3	775	14	BP160335	BP160335 603184463
8	716	40.0	908	14	BQ431184	BQ431184 AGNCOURT
9	664	37.1	1071	14	BQ035307	BQ035307 AGNCOURT
10	638	35.7	689	9	AJ451881	AJ451881 AJ451881
11	588	32.9	716	12	B6709510	B6709510 602688206
12	585	32.7	614	10	B6619078	B6619078 B6619078
13	576.5	32.2	744	9	AJ448138	AJ448138 AJ448138
14	572	32.0	526	17	AQ422321	AQ422321 RPT-11-1
15	551	30.8	515	12	BG086312	BG086312 H3124D06-
16	541	30.2	537	13	B1789500	B1789500 IC95e08-Y
17	518.5	29.0	643	10	BB628807	BB628807 BB628807
18	490	27.4	513	12	BE755721	BE755721 209634 MA
19	487.5	27.2	688	13	BJ026619	BJ026619 BJ026619
20	475	26.6	446	10	BB849603	BB849603 BB849603
21	472	26.4	404	9	AL702256	AL702256 DKFZP686H
22	449.5	25.1	540	13	BJ010671	BJ010671 BJ010671
23	444	24.8	567	13	BJ012356	BJ012356 BJ012356
24	444	24.5	1023	12	BG120825	BG120825 60348589
25	401	22.4	689	12	BG190041	BG190041 R59097 A
26	394.5	22.1	713	12	BJ024478	BJ024478 BJ024478
27	385	21.5	713	12	BJ194151	BJ194151 R512393
28	369	20.6	798	12	BG202840	BG202840 R512393
29	352	19.7	581	13	BM129591	BM129591 B112607-Y
30	350	19.6	627	10	BP128943	BP128943 B112607-Y
31	347	19.4	719	12	BP194632	BP194632 R517185
32	347	18.8	304	10	AM838362	AM838362 R517185
33	336	18.6	851	12	BM135152	BM135152 R517185
34	332	18.6	851	12	BM128641	BM128641 R517185
35	316	17.7	554	13	AA277912	AA277912 vc08c02.X
36	298	16.7	473	9	AI980778	AI980778 pat.pk003
37	288	16.1	358	10	BM487537	BM487537 pgm2n.pk0
38	273.5	15.3	358	13	AM336839	AM336839 21339 MAR
39	273	15.3	358	10	BE936477	BE936477 RCI-NT003
40	272.5	15.2	212	12	BM977108	BM977108 602846207
41	272	14.2	360	13	BM489249	BM489249 pgm2n.pk0
42	254.5	14.2	802	9	AJ394332	AJ394332 AJ394332
43	254.5	14.2	802	9	AJ394332	AJ394332 AJ394332
44	245	13.7	720	17	AZ284961	AZ284961 RPT-23-1
45	238	13.3	766	13	BI857018	BI857018 603363570

# ALIGNMENTS

RESULT 1										
B0715165										
LOCUS										
DEFINITION	B0715165	1014 bp	mRNA	linear						EST 16-JUL-2002
ACCESSION	AGNCOCOURT_8480637	NIH_MGC_129	Mus musculus	CDNA clone	IMAGE:6310003					
VERSION	B0715165									
KEYWORDS	B0715165.1	GI:21854064								
SOURCE	EST.									
ORGANISM	house mouse.									
	Mus musculus									
	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;									
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.									
	1 (bases 1 to 1014)									
REFERENCE	NIH-MGC	http://mgc.ncl.nih.gov/.								
AUTHORS	National Institutes of Health.	Mammalian Gene Collection (MGC)								
TITLE										

JOURNAL  
COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: Susan L. Sullivan, PhD.  
CDNA Library Preparation: ResGen, Invitrogen Corp.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:  
http://image.llnl.gov  
Plate: LLM13731 row: j column: 20  
High quality sequence stop: 566.  
Location/Qualifiers

FEATURES  
source

1. 1014  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="IMAGE:6310003"  
/lab\_host="NIH-MGC-129"  
/note="Organ: olfactory epithelium; Vector: pcwv-sport6.1.cdb; Site\_1: EcoRV; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2.2 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH-MGC library."  
BASE COUNT 238 a 266 c 303 g 207 t

## Alignment Scores:

Pred. No.: 3,73e-105 Length: 1014  
Score: 969.00 Matches: 176  
Percent Similarity: 89.95% Conservative: 21  
Best Local Similarity: 80.37% Mismatches: 20  
Query Match: 54.16% Indels: 2  
DB: 14 Gaps: 1

US-09-748-451-2 (1-338) x BQ751565 (1-1014)

QY 120 AAlAspLeuEntHsAnVAlSerGlnAsnHlAlAlaGluThrArgAlaGluAspPro 139

DB 1 GCGGATCTTCGTGATACGTCGACGACAGAAATATCTGCAGAGACCCGCGAGACGACCA 60

QY 140 ProTrpPheGluGluGluSerArgPheGluSerGlySerGlyTyrLeuArgTyrSer 159

DB 61 TCCTGTTTAAAGCTTTGGAGTCGAGATTCAGGAATTAAGTCGGCTATCTGAGTACAC 120

QY 160 CysGluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGly 179

DB 121 TGTGAGAGTGGATCCGGGGTACCTTAAGAGAGGTGAGCGCTTACCTCATGTGCGAT 180

QY 180 AlAGluAlaGluGluGluPheLeuArgValLeuGlySerMetCysGlnArgLeuArgSer 199

DB 181 GAAGCACTCAAGAGAGTACGTCGAGCTCTGGCTCCATGTCACAGAGCTCAAAATCG 240

QY 200 MetGlnTyrAsnGlySerTyrPheAspArgGlyAlaValGlyGlySerArgLeuGlyThr 219

DB 241 GTGCGATACAAATGCGACCTTTTCGACAGAGGTGAGAGCCACAGCCCTCTGACT 300

QY 220 ProGluGlyTyrPheSerCysGlnGlyTyrPheAspMetAspSerCysLeuSerArgHis 239

DB 301 CCAAGAGAGTGTCTCTCCAGAGCCCTTGTACCTCGAGAGCTCTTCCAAACAC 360

QY 240 SerIleAsnProTyrSerAsnArgGluSerArgIleLeuPheSerThrTrpAsnLeuAsp 259

DB 361 TCCATCAACCCCTATGCAACAGAGAGCCGATCTCTTCAGTACCTGGAACCTGGAT 420

QY 260 HisIleIleGluLysLysArgThrIleIleProThrLeuValGluAlaIleLysGluGln 279

DB 421 CAAATATATAGAGAGAGCCACCGGTGATCCACAGCTGGCTGAAGCCATC-----CAG 474

QY 280 AspGlyArgGluValAspTrpGlyTyrPheTyrGlyLeuLeuPheThrSerGluAsnLeu 299

DB 475 GATGGGAGGAGGTGAAGTGGAGTACTTCTACAGCCTGCTTCACTGCGAGAACTCG 534

QY 300 LysLeuValHisIleValCysHisLysLysThrThrHisLysLeuAsnCysAspProSer 319

DB 535 AAGTGTGTGACATCGCTGCGACAGAAACACACACAGAGTGGAGTCCGACCGCAGT 594

QY 320 ArgIleTyrAspProGlnTrpArgLeuLysArgGlyGlnProValArgLysArgGln 338

DB 595 AGGATCTATGCGCTCAGACAGATCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 651

RESULT 2 796 bp mRNA linear EST 12-SEP-2001

BI657172 603280838F1 NCL\_CGAP\_Mam4 Mus musculus CDNA clone IMAGE:5325440-3

LOCUS 11n1

DEFINITION mRNA sequence.

ACCESSION BI657172

VERSION BI657172.1 GI:15571408

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS NIH-MGC http://mgi.ncl.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapds-remail.nih.gov  
Tissue Procurement: Lotmar Hennighausen Ph.D., Priscilla Furth Ph.D.

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LLM1825 row: k column: 09

High quality sequence stop: 762.

Location/Qualifiers

1. 796

/organism="Mus musculus"

/strain="NMRI"

/db\_xref="taxon:10090"

/clone="IMAGE:5325440"

/lab\_host="NIH-MGC-Mam4"

/issue\_type="tumor, gross tissue"

/dev\_stage="5 months"

/lab\_host="DH108"

/note="Organ: mammary; Vector: pcwv-sport6; Site\_1: SalI; Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lotmar Hennighausen/Priscilla Furth, NIH Reference for transgenic model: Li et al., Cell Growth and Differentiation 7, 3-11 (1996)."

BASE COUNT 172 a 226 c 235 g 163 t

FEATURES  
source

US-09-748-451-2 (1-338) x BI657172 (1-796)

## Alignment Scores:

Pred. No.: 7.01e-104 Length: 796  
Score: 957.00 Matches: 185  
Percent Similarity: 86.29% Conservative: 29  
Best Local Similarity: 74.60% Mismatches: 34  
Query Match: 53.49% Indels: 2  
DB: 13 Gaps: 0

US-09-748-451-2 (1-338) x BI657172 (1-796)

QY 1 MetLeuGlnLysProLysSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20

DB 55 GTGCTCGCCCAACCCAAATGCGTCAAGTTCGAGCCCTACATAGCGCTCGAAGTGGC 114

QY 21 ValAlaGlyArgSerCysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuPro 40

DB 115 GTGGCGCCCGAGAGCTCCAGAGCTCTCGTAAAGGCTGCGTCCGCTTCAGACTCCG 174

OY		41	GUAAGGYSERATgLeucySAleuTYrGUAspGLYTrHguLeuPTrHguLeuAspTYRPe	60
Db		175	ANGCCCGGTTCGGCGTGCCTGTAACGAAGTAGCACCAGGAGGACGACGACGTCTTC	234
OY		61	ProSerValIProAspAsnAlaGlueuValleuLeuThrLeuGIgnAlaTrpGlnGly	80
Db		235	CCGGGCCCTCCCAACAACACGCTAGCTCCATTGTGTCACCGCTGGCACACTGCGATGGC	294
OY		81	TYrValSerAspIleArgAryPhelLeuSerAlaPheHisGluProGlnValGlyLeuIle	100
Db		255	TATGTGAGTGACATACACAGCTTTCCCTCACTGTGTATTATGAGCCATCGCCGCGTCACTC	354
OY		101	GlnAlaAlaGlnInLeuLeuCYsaSpGLuGlnAlaProGlnArgGlnArgLeuAla	120
Db		355	CAGGCTGCACGGCAACTGCTGTCACTAGACGAGGCCCCACCTGAGGCAAAGAAGCTGGTGGCC	414
OY		121	AspLeuLeuHisAsnValSerGlnAsnIleAlaAlaGlnIurHraArgAlaGluAspPropo	140
Db		415	GATCTCTGCATCTCAGTGCAGCCAGAAATATTACTGAGAGACCCGGGAGACGA-CCATCC	473
OY		141	TRPHeGluGlyLeuGluSerArGPheGlnSerLYSSerGlyTYrLeuArgTYRSerCYs	160
Db		477	TGCTTTGAAGGTTTGGAGTCTGAGATTCAAGATAAATGCTGGCTCATCGAGATACACCTCT	533
OY		161	GluSerArgIleAspSerTYrLeuArgGluValSerSerTYrProSerTrpValGlyAla	180
Db		534	GAGACTCGGATCCGGGGTTCCTTAAGAGGGAGGACCGCTTACACTCTATAGGATGA	593
OY		181	GluAlaGlnGluGluPheLeuArgValleuGlySerMetCYSGlnArgLeuArgSerMet	200
Db		594	GCAGCTCAAGAAAGTAGTACCTCGGAGCTCTTGCTTCATGCTCCAGAAAGCTCAAAATGGGG	653
OY		201	GlnTYrAnSnglySerTYrPheAspArgGlyAlaLYsgLyGlySerArgLeuCYsrTrpPro	220
Db		654	CAGTCAATATGGCACACTTATTTCGACAGAGSTGCAGAACGCCACACGCCGCTC-TGTACTCCA	712
OY		221	GluGlyTYrPheSerCYsGlnGlyProPheAspMetAspSerCYsLeuSerArgHisSer	240
Db		713	GAAAGATGCTTCTCTCTCCACGGGCCCTTTGACTCGAGAGCTGTCTTTCCAAACACTCC	772
OY		241	IleAsnProTYrSerAsnArgGlu	248
Db		773	ATCAAAACCTATGGCCACAGAGAG	796
RESULT 3	BG863734			
LOCUS	BG863734	834 bp	mRNA	linear
DEFINITION	602796717p1 NCI-CGAP Mam4 Mus musculus CDNA clone IMAGE:431772 5'			EST 26-MAY-2001
VERSION	BG863734			
KEYWORDS	BG863734.1 GI:14214272			
SOURCE	EST.			
ORGANISM	Mus musculus.			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
TITLE	NIH-MGC http://mgc.nci.nih.gov/			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999) Contact: Robert Straube, Ph.D. Email: csepds-remail.nih.gov Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth Ph.D. cDNA Library Preparation: Life Technologies, Inc. DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution Information can be found through the I.M.A.G.E. Consortium/LNL at: http://Image.lnl.gov Plate: LLM10830 Row: A Column: 05 High quality sequence stop: 788. Location/Qualifiers			

```

source                               1..834
                                     /organism="Mus musculus"
                                     /strain="NMRI"
                                     /db_xref="taxon:10090"
                                     /clone="IMAGE:6917772"
                                     /clone_1lb="NCI-GCAP_Mam4"
                                     /tissue_type="tumor_gross_tissue"
                                     /dev_stage="5 months"
                                     /lab_host="DH10B"
                                     note="Organ; Mammary; Vector: pCMV-SORT6; Site:1: SalI;
                                     Site-2: NotI; Cloned unidirectionally. Primer: Oligo dT.
                                     Library constructed by Life Technologies. Investigators
                                     providing samples: Lohar Henningshausen/Piscilla Futhr,
                                     NIH Reference for transgenic model: Li et al., Cell Growth
                                     and Differentiation 7, 3-11 (1996)."
```

BASE COUNT 186 a 236 c 240 g 172 t

ORIGIN
//ab038247.1

```

Db      707 CCAGACGAGTGTCTCTCCAGGCGCCCTTGACCTGAGAGCTGCTTCCAGGC 766
OY      239 ISSerIleAsnProTyrSerAsnArg-GluSerArgIleuPheSerThrTyrPasnLeu 258
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      767 ACTCAATCAGACCCATGACGACAGAACAGAGCGGATCTCTCAGTCC-TGGACCTG 825

RESULT 4
BIS17452
LOCUS   685 bp  mRNA  linear  EST 29-AUG-2001
DEFINITION 603041725F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5162791 5'
ACCESSION BIS17452
VERSION   BIS17452.1 GI:15342244
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 685)
           NIH-MGC http://mhc.ncl.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
           Unpublished (1999)
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: Life Technologies, Inc.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.llnl.gov
           Plate: LHAM11404 row: b column: 08
           High quality sequence start: 6
           High quality sequence stop: 685.
FEATURES
     source
         1..685
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5162791"
            /clone_id="NIH_MGC_116"
            /lab_host="DH10B"
            /note="Organ: pooled colon, kidney, stomach; Vector:
            PCWV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
            source anonymous pool of 3 colons, age 26 yo male, 49 yo
            female, 71 yo male colon; 46 yo male kidney, and pool of 2
            stomachs, 62 yo male and 70 yo female. Library is
            oligo-dT primed and directionally cloned (EcoRV site is
            destroyed upon cloning). Average insert size 1.4 kb,
            insert size range 1-3 kb. Library is normalized and
            enriched for full-length clones and was constructed by C.
            Gruber (Invitrogen). Research Genetics tracking code
            023. Note: this is a NIH-MGC library."
COUNT  121 a 202 c 236 g 126 t

Alignment Scores:
Pred. No.: 1.15e-93 Length: 685
Score: 871.00 Matches: 179
Percent Similarity: 89.95% Conservative: 0
Best local Similarity: 89.95% Mismatches: 2
Query Match: 48.69% Indels: 19
DB: 13 Gaps: 1

US-09-748-451-2 (1-338) x BIS17452 (1-685)
OY      1 MetLeuGlnIysProTyrSerValLysLeuArgAlaLeuArgSerProArgLysPheGly 20
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      91 ATGCTCCAGAGGCCCAAGAGCGTGAAGCTGGCGCCCTGCGCAGCCGAGGAAGTTCCGC 150
OY      21 ValAlaGlyArgSerCysGlnGluValLeuArg-LysGlyCysLeuArgPheGlnLeuPr 40
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      151 GTGCTGCGCGGAGCAGCTCTGCTCTGCAAGGCGCTGTCTCCGCTTCCAGCTCC 210

```

```

OY      40 oGluArgGlySerArgLeuCysLeuTyrGluAspGlyThrGluLeuThrGluAspTyrPh 60
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      211 TGAGCGCGGTGCCGGCTGCTGCTGTACGAGGATGCGACGGAGCTGACGGAATTTACTT 270
OY      60 eProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaTrrPglngl 80
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      271 CCCAGTGTCCCGCAACAGCCGAGCTGTCTCTCCTCCTCCTGAGGCCAGGCTGGCAGGG 330
OY      80 YTr------ -ValSe 83
        |
Db      331 CTG-ATGTCCTTCTGCTGACCGGACCTTTTGTCTTCCATTGGTCGAGATGTGAG 389
OY      83 rAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaI 103
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      390 CGACATCAGCGGCTTCTCTCAGTGCATTCACGAGCAGACAGTGGGGCTCATCCAGCGCGC 449
OY      103 aGlnGlnLeuLeuCysAspGluGlnAlaProGlnArgGlnArgLeuLeuAlaAspLeuLe 123
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      450 CCACGAGCTGTGTGTGTGAGCAGCGCCACAGAGGAGGCTGCTGCTGACCTCCCT 509
OY      123 uHisAsnValSerGlnAsnIleAlaAlaGluThrArgAlaGluAspPropTrrPheG 143
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      510 GCACAGCTCAGCCAGAACATCGCGCGGAGACCCGGCTGAGACCCCGCTGTTCGA 569
OY      143 uGlyLeuGluSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerAr 163
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      570 AGCTGTGAGGCCGATTCAGACAGTGTGCTATCTGATGATACAGCTGTGAGAGCGC 629
OY      163 gIleArgSerTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGlu 181
        |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db      630 GATCCGGAGTTACTGAGGAGGTGAGCTCTTACCCTCCAGCGGTGGGTGGCGAG 684

RESULT 5
B1770511
LOCUS   830 bp  mRNA  linear  EST 25-SEP-2001
DEFINITION 603055488F1 NIH_MGC_122 Homo sapiens cDNA clone IMAGE:5204996 5'
ACCESSION B1770511
VERSION   B1770511.1 GI:15762089
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 830)
           NIH-MGC http://mhc.ncl.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
           Unpublished (1999)
           Contact: Robert Strausberg, Ph.D.
           Email: cgapbs-remail.nih.gov
           Tissue Procurement: Life Technologies, Inc.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/LNL at:
           http://image.llnl.gov
           Plate: LHAM11513 row: p column: 21
           High quality sequence stop: 820.
FEATURES
     source
         1..830
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:5204996"
            /clone_id="NIH_MGC_122"
            /lab_host="DH10B"
            /note="Organ: pooled lung and spleen; Vector: PCWV-SPORT6;
            Site_1: NotI; Site_2: EcoRV (destroyed); RNA source
            anonymous pool of 24 week female lung, 16 week female
            spleen, and 20-22 week male spleens. Library is oligo-dT
            primed and directionally cloned (EcoRV site is destroyed
            upon cloning). Average insert size 1.4 kb, insert size

```

range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 026. Note: this is a NIH-MGC Library."

BASE COUNT 145 a 249 c 280 g 155 t 1 others

## ALIGNMENT SCORES:

Pred. No.: 6,81e-82 Length: 830  
Score: 774.00 Matches: 160  
Percent Similarity: 94.67% Conservative: 0  
Best Local Similarity: 94.67% Mismatches: 8  
Query Match: 43.26% Indels: 3  
DB: 13 Gaps: 0

-748-451-2 (1-338) x B1770511 (1-830)

```

1 MetLeuGlnlyProlySerVallyLeuArgAlaLeuArgSerProArglyPhegly 20
  |||||||
324 ATGCTCAAGAGCCCAAGAGCGTGAAGCTGGGCGCCGCGAGCCGAGAGTTCCGCG 383
  |||||||
21 ValAlaGlyArgSerCysGlnGluValLeuArglyGlyCysLeuArgPheGlnLeuPro 40
  |||||||
384 GTGCTGGCGCGAGCTGCCAGAGAGTCTCCGCAAGGCGTCTCCGCTCCAGCTCCCT 443
  |||||||
41 GluArgGlySerArgLeuCysLeuTyrluAspGlyThrcyluLeuThrcyluAspTyrPhe 60
  |||||||
444 GAGCGCGGTTCCCGGCTGCTGCTGACGAGATGCGCAGGAGCTGACGAGAGATTACTTC 503
  |||||||
61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrcyluGlnAlaTrpGlnly 80
  |||||||
504 CCCAGTCTCCCGAGACAGCGAGCTGGTGTCTACCTGGGCGAGGCGGAGGCG 563
  |||||||
81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeu 100
  |||||||
564 TATGTAGCGACATGAGCGCTGCTCTCAGTTCAGTTCAGCAGCAGCAGAGTGGGCTCATC 623
  |||||||
101 GlnAlaAlaGlnGlnLeuLeuCysAspGlnAlaProGlnArgGlnArgPheLeuAla 120
  |||||||
624 CAGCCGCGCAGCAGCTGCTGCTGATGAGAGAGGCCGCCAGAGGAGAGGCTGCTGCT 683
  |||||||
121 AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrcyluArgAlaAspPro 140
  |||||||
684 GACTC-CTGCACACAGCTCAGCAGACATGCGCTGGCAGACCCGGCTGAGCCGAG 742
  |||||||
141 TrpPheGlnGlyLeuGlu-SerArgPheGlnSerLySerGlyTyrLeuArgTyrSer 160
  |||||||
743 GG-TTTGAAGGCTTGAAGTCCGATTTCAGAGCAGCATGCTATCTGATGATACAGCTG 801
  |||||||
160 SglSerArgIleArgSerTyrLeu 168
  |||||||
802 TGAGAGCCGAGATCCGGAGATTACTG 826
  |||||||

```

RESULT 6  
B165515  
LOCUS 855 bp mRNA linear **EST 30-OCT-2000**  
DEFINITION 601777413F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:4019043-5  
mRNA sequence.  
ACCESSION B165515  
VERSION B165515.1 GI:11045867  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 855)  
NIH-MGC <http://mgs.nci.nih.gov/>  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: [cgap@nci.nih.gov](mailto:cgap@nci.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.

## FEATURES

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNLN at:  
<http://image.llnl.gov>  
Plate: LLM9270 row: j column: 04  
High quality sequence stop: 683.  
Location/Qualifiers  
1..855  
/organism="Mus musculus"  
/strain="CZECH 11"  
/db\_xref="taxon:10090"  
/clone="IMAGE:4019043"  
/clone\_1lb="NCI\_CGAP\_Lu29"  
/tissue\_type="spontaneous tumor, metastatic to mammary."  
Stem Cell Origin  
/lab\_host="DH10B"  
/note="Organ: Lung; Vector: pCMV-SPORT6; Site\_1: Salt;  
Site\_2: Not; Cloned unidirectionally. Primer: oligo dt.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

BASE COUNT 180 a 215 c 285 g 174 t 1 others

## ALIGNMENT SCORES:

Pred. No.: 2.59e-80 Length: 855  
Score: 761.00 Matches: 160  
Percent Similarity: 79.41% Conservative: 29  
Best Local Similarity: 67.23% Mismatches: 44  
Query Match: 42.54% Indels: 7  
DB: 12 Gaps: 1

US-09-748-451-2 (1-338) x B165515 (1-855)

```

1 MetLeuGlnlyProlySerVallyLeuArgAlaLeuArgSerProArglyPhegly 20
  |||||||
62 GTGCTCCGCAACCCAAATGCTGAAGTTGGAGCCCTACATAGCCGCTCGAAGTTCCGCG 121
  |||||||
21 ValAlaGlyArgSerCysGlnGluValLeuArglyGlyCysLeuArgPheGlnLeuPro 40
  |||||||
122 GTGCGCGCCCGAGCTGCCAGAGCTGCTGCTGAAGGCTGCTGCTCCGCTCCAGTCCCG 181
  |||||||
41 GluArgGlySerArgLeuCysLeuTyrluAspGlyThrcyluLeuThrcyluAspTyrPhe 60
  |||||||
182 ATGCGCGTCCCGGCTGCTGCTGACAGATGCGCAGGAGTGCAGCAGCAGCTGCTTC 241
  |||||||
61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrcyluGlnAlaTrpGlnly 80
  |||||||
242 CCGGCGCTTCCCAACAGCAGCTGAGCTCTATGCTCACCGCTGGCAGACCTGGCATGCG 301
  |||||||
81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeu 100
  |||||||
302 TATGTAGTGACATGACACAGCTTCCCTCAGTGTGTTAATGAGCCACATGCGCGCTCATC 361
  |||||||
101 GlnAlaAlaGlnGlnLeuLeuCysAspGlnAlaProGlnArgGlnArgPheLeuAla 120
  |||||||
362 CAGGCTCAGCGCAACTGCTGTCAGATGAGAGGCCGCCACTGAGGAGCAAGCTGCTGCC 421
  |||||||
121 AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrcyluArgAlaAspPro 140
  |||||||
422 GATCTTGTGATCAGTGCAGCAGAAATTTCTGTCAGAGACCCGGGAGAGAGACCATCC 481
  |||||||
141 TrpPheGlnGlyLeuGlu-SerArgPheGlnSerLySerGlyTyrLeuArgTyrSer 160
  |||||||
482 TCGTTTAAGCTTG-CAGTCAGATTCAGAGAAATAGCGGCTATCTGATGATACAGCTG 540
  |||||||
161 GluSerArgIleArgSerTyrLeuArgGluValSerSerTyrProSerThyAlaGlyAla 180
  |||||||
541 GAGAGTGGATCCGGGCTTACTTAAAGAGTGGAGCCCTTACCTCTTATGTTGGATGAGAA 600
  |||||||
181 GluAlaGlnGlnLeuLeu-ArgValLeuGlySerMetCysGlnArgLeuArgSerMe 200
  |||||||
601 GCAGTGCAGAGAGTACTGTCAGAGTCTTGAGTTCATGTGAG-AAAGCTCAATTCGGT 659
  |||||||

```

QY	200	TGATTTAASAGLYSerIYrheAspaRgVGLyAlAALyGLYSerARgLeucysThr--	219
	:		:
Dd	660	GCACTACAAATAATGGCAGGTATTTCACACAGAGTGCGGAAGAACCAGACGCCCTTCTGTCCAG	719
QY	220	ProGUglYTrPpheSerCySginglyPrIoPheAspMetaspsercys	235
	:		:
Dd	720	AAGGAGGGT-----TCCTGGCAGGGGCCCTTTCACCTCGGAGAGGGGTGT	761
RESULT 7			
BI916035/c			
LOCUS			
DEFINITION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
SEQUENCE			
CHORS			
FILE			
JOURNAL			
COMMENT			

```

Plate: L1AM11627 row: m column: 15
High quality sequence stop: 709.
FEATURES
    source          location/Qualifiers
    1. . 775

```

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_image="5248694"
/clone_id="N1H_MGC_121"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (IMMUTAGEN). Research Genetics Tracking code 017..Note:

```

COUNT	151 a	233 c	239 g	152 t
-------	-------	-------	-------	-------

Alignment Scores:	
Pred. No.:	1,28e-77
Score:	778.00
Percent Similarity:	88.00%
Best Local Similarity:	81.00%
Query Match:	41.25%
OB:	13
Length:	775
Matches:	1675
Conservative:	14
Mismatches:	20
Indels:	6
Gaps:	0

US-09-748-451-2 (1-338) x BI916035 (1-775)

QY	6	SeVtAlPrOaSPnAlaIuLeuValLeuLeuTrILeuGlyGlnAlaTrp-GlnGlyTY	81
Db	740	AGGTCCTCGAGAGATCCGAGCTGGTCCGCTCACCTCAAGCCAGGCTTGGCCAGGGCTA	688
QY	81	tVtAlSerAlIeAtgArgPhoLeuSerAlaPheHisGluProGlnValGlyLeuIleG	107
Db	680	TGGATGACAT-CGGTGGCTTCTAGTGCCTTTTGGCCAGCGGACGGGCTTCATCA	622

QY	101	n-fAlaIaGInGInLeuLeuCySAspGInGInAla-ProGInAArgInAArgLeuLeuAla	120
Db	621	AGGCCCCGGGCACTGCTGTATGTACAGAGCCCCCAGAGGAGCAAGACTGTGGCC	562
QY	121	AspLeuLeuHisAsnValSerGlnAsnIleAlaGluThrArgAlaGluAspPro	140
Db	561	GACCTCTGTACAACTCAGCAGCAAGCTCGAGCCGACACCCGGCTGAGAGCCGGCTG	502
QY	141	TrpPhGInGInLeuGInSerArpPhGInSerIYssSerGInYTrIleuArGTrYrSerCys	160
Db	501	TGCTTTGAAGCTTGAGATCCCGATTTTGGATTAAGTGTGGCTAATCTGAATACAGCTG	442
QY	161	GluSerArgIleArGserTrYrLeuArgGInuValSerTrYrProSerThrValGlyAla	180
Db	441	GAAGCCGGATCCGGATCTGAGAGAGGTGAGCTCTGCTCCATCAGTGGGGTCCA	382
QY	181	GluAlaGInGInGInuPhLeuArGValLeuGlySerMetCysGlnArGLeuArGSerMet	200
Db	381	GAGGCTCAGAGAGAAATCTCTGGGGTCTCTCGCTCCATGTCCTCCAGAAAGCTTCAGTGTG	322
QY	201	GInTrYAsnGlySerTrYrPheAspArgGlyAlaIYssGlySerArgLeuCyThrPro	220
Db	321	CATATAGAGGCGCAGCATATAC-AACAGAGAGCCAAAGGTGGCAACCCCTCTGTGCACACCG	263
QY	221	GluGlyTrpPheSerCysGInGlyProDPhAspMetAspSerCysLeuSerArgHisSer	240
Db	262	GAAGGCTAATGTCCTCCGACGAGGTCCTTCCACATGACGAGCGTCTTGTCAAAATACTCC	203
QY	241	IleAsnProTrYSerAsn-ArgGInuSerArgIleLeuPheSerThrTrpAsn	257
Db	202	ATACACCCCTTACAGTACACGGGAGAGAGAGCTCTCTTCTTGAGACTGGAAAC	151

RESULT 8					
BO431184					
LOCUS					
DEFINITION	BO431184	908 bp	mRNA	linear	EST 24-MAY-2001
	AGENOOUT-7786920	NIH_MGC_92	Homo sapiens	CDNA clone	IMAGE:5058785
	5', mRNA sequence.				

ACCESSION VERSION KEYWORDS SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
BQ431184 BQ431184.1 GI:21170260 EST.	human.	Homo sapiens	Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	1 (bases 1 to 908) NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a> .	National Institutes of Health, Mammalian Gene Collection (MGC)	Unpublished (1999) Contact: Robert Strausberg, Ph.D.

**FEATURES**  
**SOURCE**

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6068786"
/clone_lib="NIH-MGC_92"
/tissue-type="embryonic carcinoma, cell line"
/lab_host="DH10b (phage-resistant)"
/note="Organ: testis; Vector: PCMV-SPORE6; Site_1: NOT1; Site_2: Salt. Cloned unidirectionally; oligo-primed. Average insert size 2.5 kb. Library enriched for full-length clones and constructed by Life Technologies. Note: this is a NIH-MGC library."

```

BASE COUNT 158 a 271 c 307 g 171 t 1 others  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 7.15e-75 Length: 908  
 Score: 716.00 Matches: 169  
 Percent Similarity: 87.318 Conservative: 3  
 Best Local Similarity: 85.798 Mismatches: 13  
 Query Match: 40.028 Indels: 12  
 DB: 14 Gaps: 2

US-09-748-451-2 (1-338) x BQ431184 (1-908)

QY 1 MetLeuGlnIysProlYsSerValIysLeuArgAlaLeuArgSerProArgIysPheGly 20  
 |||||  
 312 ATGCTCCAGAACCCCAAGAGCTGAAAGCTGGCGCCCTCGCGCAGCCGAGAGAGTTGCGC 371  
 21 ValAlaGlyArgSerCysGlnGlnValLeuArgIysGlyCysLeuArgPheGlnLeuPro 40  
 |||||  
 372 GTGGCTGGCCGGAGCTGCCAGAGAGTGGTGGCAAGGGGCTGTCCGCTCCAGCTCCCT 431  
 QY 41 GluArgIysSerArgIysLeuArgIysLeuArgIysLeuArgIysLeuArgIysLeu 60  
 |||||  
 432 GAGCGCGGTTCCCGGCTGTGCTGTACAGAGTGGCAGAGCTGACGAGAGATTACTTC 491  
 QY 61 ProSerValProAspAsnAlaGluLeuValLeuLeuThrLeuGlnAlaArgPheGln 80  
 |||||  
 492 CCGAGTCTCCGACACACAGCCGAGCTGGTGGCTGCTGCTGAGCCAGGCGCTGGCGGCG 551  
 QY 81 TyrValSerAspIleArgArgPheLeuSerAla-PheHisGlnProGlnValGlyLeu 100  
 |||||  
 552 TATGTGAGCAGACATCAGGGGCTTCTCAGTGCATTTTCAGAGAGCCGAGGGGCTCAT 611  
 QY 100 eGlnAla-AlaGlnGlnLeuLeuCysAspGlnGlnAlaProGlnArgGln-ArgLeuLeu 119  
 |||||  
 612 CCAGGCCCCCGCAGCAGCTGCTGTGTGATGAGAGCCGCGCAGAGAGAGAGAGAGAGAG 671  
 QY 120 AlaAspLeuLeuHisAsnValSerGlnAsnIleAla-AlaGlnThrArgAlaGluAsp 139  
 |||||  
 672 GCTGACCTCTCCGACACAGCTGAGCAACATCCGCGGCGAGAGCCGCGGCTGGAGACCC 731  
 QY 139 oProTyr-PheGlnGlyLeuGlu---SerArgPheGlnSerIysSer-GlyTyrLeu-Arg 157  
 |||||  
 732 CGCGTGGTTTGAAGGGCTGGAAAGCCGCAATTCAGAGAGAGTCTGGCTATCTGAGAG 791  
 QY 157 gTyr-SerCysGlnSerArgIysArgSer-TyrLeuArg-GluValSerSerTyr---Pr 175  
 |||||  
 792 ATACACTGTGAAACCGGATCCGGATTACCTGAGAGGAGAGTGAAGCTTCTACCCCC 851  
 QY 175 cSerThrValGlyAlaGlnAlaGlnGlu 185  
 |||||  
 852 TCCAACGGTGGTTCGCCGAAGGCTTCACAAA 882

DB 852 TCCAACGGTGGTTCGCCGAAGGCTTCACAAA 882

RESULT 9  
 BQ05307/c 1071 bp mRNA linear [EST-29-MAR-2002]  
 LOCUS AGENCOURT.6797380 NIH\_MGC\_99 Homo sapiens cDNA clone IMAGE:5807463  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BQ05307  
 VERSION BQ05307.1 GI:19814647  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 1071)  
 AUTHORS NIH-MGC <http://mhc.nci.nih.gov/>.  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: Lou Staudt  
 cDNA Library Preparation: Rubin Laboratory

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LUCM2049 row: 0 column: 16  
 High quality sequence stop: 595.  
 Location/Qualifiers  
 1. 1071  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5807463"  
 /clone\_lib="NIH\_MGC\_99"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: Lymph. Vector: pORF7; Site:1: XhoI; Site:2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
 into EcoRI/XhoI sites using the following 5' adaptor:  
 GGCAAGG(G). Size-selected >500bp for average insert size  
 1.8kb. Library constructed by Ling Hong in the laboratory  
 of Gerald M. Rubin (University of California, Berkeley)  
 using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
 II RT (Life Technologies). Note: this is a NIH\_MGC  
 Library."  
 BASE COUNT 248 a 344 c 278 g 201 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.61e-68 Length: 1071  
 Score: 664.00 Matches: 145  
 Percent Similarity: 83.77% Conservative: 15  
 Best Local Similarity: 75.92% Mismatches: 19  
 Query Match: 37.12% Indels: 13  
 DB: 14 Gaps: 2

US-09-748-451-2 (1-338) x BQ05307 (1-1071)

QY 76 GlnAlaArgPheGlnGlyTyrVal-----SerAspIleArgArgPheLeuSer 90  
 |||||  
 921 AGGCGTGGCGAGGGCTTATTGAGAGACATCCGCGGCTTCTCAGGTGCTT----- 868  
 QY 91 AlaPheHisGluProGlnValGlyLeuIleGlnAla-AlaGlnGlnLeuLeuCysAsp 110  
 |||||  
 867 ---TTCCGCAAGCGGAGAGGGGGTTCATCCAAAGCCGCGGCAATTAAGTGTGATGA 811  
 QY 110 uGlnAla-ProGlnArgGlnArgIleuLeuAlaAspLeuLeuHisAsnValSerGlnAsn 130  
 |||||  
 810 GCAGGCCCGCCAGAGGACAGAGACTGCTGCGCACTCTCTACAGAGTACAGCAAGAG 751  
 QY 130 le-AlaAlaGlnThrArgAlaGluAspProPro-TyrPheGlnGlyLeuGlnSerArg 149  
 |||||  
 750 TCCGAGCGCCAGACCCGCGCTGAGACCCCGCTGGTGAAGCTTGAAGTCCCGATT 691  
 QY 149 eGlnSerIysSerGlyTyrLeuArgTyrSerCysGlnSerArgIleArgSerTyrLeu 169  
 |||||  
 690 TTGGAAATAGCTGTGCTATCTGAGATACAGCTGTGAAGCCGAGTCCGAGTACCTGAG 631  
 QY 169 gGluValSerSerTyrProSerThrValGlyAlaGlnAlaGlnGluGlnValPheLeuArg 189  
 |||||  
 630 AGAGGTGAGCTCTGCTCTCTCATGTGTGGTGCAGAGGCTGAGAGGAATTCCTCGGGGT 571  
 QY 189 IleuGlySerMetCysGlnArgLeuArgIysSerMetGlnTyrAsnGlySerTyrPheAsp 209  
 |||||  
 570 CTTCCGCTCCATGTGGCCAGAGCTCAGCTGTGAGTGAAGCAGCGCAGCTATAC-AACG 512  
 QY 209 gGlyAlaValGlyIysSerArgPheLeuArgIysThrProGlnGlyTyrPheSerCysGln 229  
 |||||  
 511 AGGAGCCAAAGGTGGAGCGGCTGTGACACCCGGAAGAGTCTCTCGCCAGGGTCC 452  
 QY 229 oPheAspMetAspSerCysLeuSerArgIysSerIleAsnProArgIysSerAsnArgGlu 249  
 |||||  
 451 CTTCGACATGACGCGTGTCTTGCAGAGATACCTCATCAACCCCTACAGTACAGGAGAG 392

OY 249 TArglleuPheSerThrTTPasn 257  
 DB 391 CAGGCTCCTCTTGAGCTGGAGAC 367  
 RESULT 10  
 LOCUS AJ451681 689 bp mRNA linear EST 22-APR-2002  
 DEFINITION AJ451681 riken1 Gallus gallus cdna clone 29j11r1, mRNA sequence.  
 ACCESSION AJ451681 GI:20261777  
 VERSION AJ451681.1  
 KEYWORDS EST.  
 SOURCE chicken.  
 ORGANISM Gallus gallus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;  
 Phasianinae; Gallus.  
 1 (bases 1 to 689)  
 Buerstedde, J. M.  
 Gallus gallus bursal lymphocyte EST  
 Unpublished (2002)  
 CONTACT: Buerstedde JM  
 Cellular Immunology  
 Heinrich-Pette-Institute  
 Martinstr. 52, 20251 Hamburg  
 Email: <http://genetics.mp1.uni-hamburg.de/dt40est.html>.  
 FEATURES  
 Source  
 1. 689  
 /organism="Gallus gallus"  
 /db\_xref="taxon:9031"  
 /clone="29j11r1"  
 /clone\_lib="riken1"  
 /cell\_type="bursal lymphocyte"  
 /dev\_stage="2-3 weeks old"  
 /note="CB inbred strain"  
 BASE COUNT 166 a 172 c 219 g 132 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 1.06e-65 Length: 689  
 Score: 638.00 Matches: 130  
 Percent Similarity: 51.60% Conservative: 31  
 Best Local Similarity: 41.67% Mismatches: 48  
 Query Match: 35.66% Indels: 103  
 DB: 9 Gaps: 2  
 US-09-748-451-2 (1-338) x AJ451681 (1-689)  
 OY 6 LysSerValLysLeuArgAlaLeuArgSerProArgLysPheGlyValAlaGlyArgSer 25  
 DB 56 AGGGGGTCCCGCTCCCGCCCGCCGCGGCGGAGAGGTTTCGGGGCGGCGCGGCGGAGC 115  
 26 CysGlnGluValLeuArgLysGlyCysLeuArgPheGlnLeuProGlnArgGlySerArg 45  
 DB 116 CTGCTGGGCTGCTGCGGAGGCGGTGCTGCGAGCTCCGCTGGGCGGAGCGCGG 175  
 OY 46 LeuGlySerLeuTyrGlnAspGlyThrGluLeuThrGluAspTyrTrpPheProSerValProAsp 65  
 DB 176 CTGTGCTCTTACGAGAGGAGCGGAGCGGAGCTGCTCCGACGCTGCGCGG 235  
 OY 66 AsnAlaGluValLeuLeuThrLeuGlyGlnAlaTrpGlnGlyTyrValSerAspIle 85  
 DB 236 CAGACGGAGCTGGCTGCTGCGGCGGCGGAGAGCTGCGCGGCGGCGGCGGCGG 280  
 OY 86 ArgArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIleGlnAlaGlnGln 105  
 DB 280 ----- 280  
 OY 106 LeuLeuGlyAspGluGlnAlaProGlnArgGlnArgLeuAlaAspLeuLeuHisAsn 125  
 DB 280 ----- 280  
 OY 126 ValSerGlnAsnIleAlaGlnGluThrArgAlaGluAspProProTrpPheGlnGluLeu 145

DB 280 ----- 280  
 OY 146 GluSerArgPheGlnSerLysSerGlyTyrLeuArgTyrSerCysGluSerArgIleArg 165  
 DB 280 ----- 280  
 OY 166 SerTyrLeuArgGluValSerSerTyrProSerThrValGlyAlaGlnAlaGlnGlu 185  
 DB 281 -----TCGAGAGATGCA 292  
 OY 186 PheLeuArgValLeuGlySerMetCysGlnArgLeuArgSerMetGlnTyrPasnLysSer 205  
 DB 293 TATAAAGGATACCTGACCTGATGTTGATTAACGAAATCTGTGAAGATTAACGCGAGC 352  
 OY 206 TyrPheAspArg---GlyAlaLysGlyLysSerArgLeuGlyCysThrProGluGlyTrpPhe 224  
 DB 353 TACTTTGACAGCAGCAGAGAGAGAGCGCGGAGTCCCTGTGCAAAAGAGAGGATGTTTC 412  
 OY 225 SerCysGlnGlyProPheAspMetAspSerCysLeuSerArgHisSerIleAsnProTyr 244  
 DB 413 TCTTGTACAGGCGCTTTTGACAGAGATGATGCTCCCATGTAAACATTTATCAACCCCTAC 472  
 OY 245 SerAsnArgGluSerArgIleLeuPheSerThrTTPasnLeuAspHisIleIleGluLys 264  
 DB 473 ACTAACAGGAGAAAGCAAGATCTCTTCAGACCTGGAATCTGTACATCAATATAGAAAG 532  
 OY 265 LysArgThrIleIleProThrLeuValGlnAlaIleLysGluGlnAspGlyArgGluVal 284  
 DB 533 AAAGCTAGAGTGTCTCCAGAACTGGCAGAGCTGCAAAACAGAGAGAGAGAGAG 592  
 OY 285 AspTrpGluTyrPheTyrGlyLeuLeuPheThrSerGluAsnLysLeuValHisIle 304  
 DB 593 AACTGGAGTACTTTATCAGCTATGTTTACAAATGAATTAATCAAACTGTACATAT 652  
 OY 305 ValCysHisLysLysThrHisLysLeuAsn 315  
 DB 653 GCTTTCACACAGAAACCAATCAATCAATCTCAGC 686  
 RESULT 11  
 BG705510 716 bp mRNA linear EST 07-MAY-2001  
 LOCUS BG705510  
 DEFINITION 602688206r1 NIH\_MGC\_95 Homo sapiens cdna clone IMAGE:4820856 5',  
 mRNA sequence.  
 BG705510  
 ACCESSION BG705510 GI:13979919  
 VERSION BG705510.1  
 KEYWORDS EST.  
 ORGANISM human.  
 SOURCE  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 716)  
 NIH-MGC <http://mhc.nci.nih.gov/>.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 CONTACT: Robert Strausberg, Ph.D.  
 Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov).  
 Tissue Procurement: Miklos Palkovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shitaki  
 Toshiyuki and Piero Carninci (RIKEN)  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (ULN)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/ULN at:  
<http://image.llnl.gov>  
 Plate: LLM10726 row: k column: 01  
 High quality sequence start: 10  
 High quality sequence stop: 573.  
 Location/Qualifiers  
 1. 716  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4820856"  
 /clone\_lib="NIH\_MGC\_95"





ACCESSION AL602837  
 VERSION AL602837.1 GI:15166343  
 KEYWORDS EST  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 637)  
 AUTHORS Wambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.  
 TITLE EST (Wambutt, R., Heubner, D., Mewes, W., Well, B. and Wiemann, S.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Wambutt, R.  
 MFS  
 Am Klopferplatz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by AGOWA (Berlin/Germany) within the cDNA sequencing consortium of the German Genome Project.  
 No sl sequence available.  
 This clone (DKFZ6860116) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 FEATURES  
 source  
 1. 637  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZ6860116"  
 /clone\_lib="686 (synonym: hlcc3)"  
 /tissue\_type="human skeletal muscle"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pT7Blue2; Site\_1: SfiI; Site\_2: SfiI; cDNA collection"  
 BASE COUNT 140 a 172 c 190 g 135 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 2,57e-58 Length: 637  
 Score: 576.00 Matches: 108  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 32.20% Indels: 0  
 DB: 9 Gaps: 0  
 US-09-748-451-2 (1-338) x AL602837 (1-637)  
 144 GATGCTGAGTCCGATTCAGAGCAAGTCTGCTATCTGAGATACAGCTGAGAGCCGG 371  
 312 GCTTGGATCCGATTCAGAGCAAGTCTGCTATCTGAGATACAGCTGAGAGCCGG 371  
 164 TleargserTleuargGluvalserSerTyrProserThraValcylAlaGluAlaGln 183  
 372 ATCCGAGATTACCTGAGAGAGGTAGCTCTACCCCTCCACGGTGGTCCGAGAGCTCAG 431  
 184 GluGluPhelauargValleuGlySerMetCysGlnarGluarGlySerMetGlnTyrAsn 203  
 432 GAGGAATTCCTCGCGGCTCTCGCCCTCATGTGCCAGAGGCTCCGCTCATGTGACATACAT 491  
 204 GlySerTyrPheAspArgGlyAlaValcylGlySerArgLeuGlyThrProGluGlyTyrP 223  
 492 GCGAGCTACTCGACGAGAGAGCCAGAGGCGCGCGCTCTGACACACCGAGAGCTGG 551  
 224 PheSerCysGlnGlyProPheAspMetAspSerCysLeuSerArgTyrGlySerIleAsnPro 243  
 552 TTCTCTGCGACAGGCTCTTGTGACATGACATGACATGACATGACATGACATGACATGACAT 611  
 244 TyrSerAsnArgGluSerArgTyr 251  
 612 TACAGTACACGAGAGAGAGAGATC 635  
 DB  
 RESULT 15  
 A0422321/c

LOCUS A0422321 526 bp DNA linear (SSS 23MAY1999)  
 DEFINITION RPCR-11-186H3.TV RPCR-11 Homo sapiens genomic clone RPCR-11-186H3.  
 ACCESSION A0422321  
 VERSION A0422321.1 GI:4480045  
 KEYWORDS GSS.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 526)  
 AUTHORS Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter, J.C.  
 TITLE Use of BAC End Sequences from Library RPCR-11 for Sequence-Ready Map Building  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Shanying Zhao, William Nierman, Mark Adams  
 Department of Eukaryotic Genomes  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: hbeet@igf.org  
 Emails are derived from the human BAC library RPCR-11. For BAC library availability, please contact Pieter de Jong (pieter@jgong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genetics (http://resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
 Seq primer: T7  
 Class: BAC ends.  
 FEATURES  
 source  
 1. 526  
 /organism="Homo sapiens"  
 /db\_xref="GDB:7571210"  
 /db\_xref="taxon:9606"  
 /clone="RPCR-11-186H3"  
 /clone\_lib="RPCR-11"  
 /sex="Male"  
 /cell\_type="Lymphocytes"  
 /note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; RPCR11 Human Male BAC library"  
 BASE COUNT 101 a 176 c 147 g 102 t  
 ORIGIN  
 Alignment Scores:  
 Pred. No.: 5,74e-58 Length: 526  
 Score: 572.00 Matches: 135  
 Percent Similarity: 78.92% Conservative: 11  
 Best Local Similarity: 72.97% Mismatches: 20  
 Query Match: 31.97% Indels: 22  
 DB: 17 Gaps: 3  
 US-09-748-451-2 (1-338) x A0422321 (1-526)  
 1 MetLeuGlnIysPheProIysSerValIysLeuArgAlaLeuArgSerProArgIysPheGly 20  
 498 GTGCTCCAGAACACACACTGTGAAGCTCGCGGCCCTGTGCAGCCCGAGAGATTGGC 439  
 21 ValAlaGlyArgSerCysGlnGluValLeuArgGlyGlyIysLeuArgPheGlnLeuPro 40  
 438 ATGCGGCGGAGAGAGTCCGCGGAGATGCTCGCAGAGGCTACCTCGCTTACAGCTCCT 379  
 41 GluArgGlySerArgLeuGlyArgIysPheGlyThrGluLeuThrGluArgPhe 60  
 378 CG-----TACAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 348  
 61 ProSerValProAspAsnAlaGluValLeuLeuThrLeuGlyAlaValArgPheGlnGly 80  
 347 -----GTCTCGACATCGCAGCTGTGTGCTGCTACCTGACCGCCAGGCTGGAGGCG 294  
 81 TyrValSerAspIleArgArgPheLeuSerAlaPheIleGlnIleValcylLeuIle 100



TITLE	JOURNAL	COMMENT
Jackson, Y. and Bowers, Y. Endocrine Pancreas Consortium unpublished (2000)		Contact: Douglas Melton, Klaus

**FEATURES.**  
**SOURCE**

COUNT	108 a	162 c	156 g	111 t
IN				

[illegible]

US-09-748-451-2 (1-338) x BI789500 (1-537)

QY MetLeuGlnIysProIySSerValIySLysLeuRgAlaLeuRgSerProRgIySPheIy 20  
 Db GTGCTCCGCCAACCCAAATGCTGCAAGTTGGAGCCCTACATGAGCCCTGCAGAAATTGGCC 39  
 QY 21 ValAlaGlyAngSerCysGlnGluValaLeuRgIySLySGIyCysLeuRgPheIleuPro 40  
 Db 99 GTGGGGGCCCGGAGCTCCAGAGAGCTGCTGCTGTAAGGGCTGCTGGCTTCAGACTCCG 159  
 QY 41 GluArgGlySerArgLeuCysLeuIyRgIuSLySpGlyThrGluLeuThrGluAspTyrPhe 60  
 Db 159 ATGCCGGATTCGGGCTGCTGCTGTAAGAAATGGCAGGATTTTATAGACAGCTTC 216

QY	61	ProSerValaIProAspAsnAlaGluLeuValleuValleuThrLeuGIAlaIArgGly	80
Db	219	CGGGCCCTCCCAACGACGCTGAGCTCATTTGCTCACCCCTGGCAGACCGCATGGCC	278
QY	81	TyrValSerAspIleArgPheLeuSerAlaPheHisGluProGlnValGlyLeuIle	100
Db	279	TATGTGATGTGCATACACAGTTCCTCAGTGTGTTAATGAGCCATATCGCGGTCATC	338
QY	101	GlnAlaAlaGlnGlnLeuLeuLeuCysAspGlnGlnAlaProGlnArgGlnLeuAla	120
Db	339	CAGCGTCACGCGCAACTGCTGTCAATGAGCAGGCCCCCATGAGGCAAAAGCTGTGGCC	398
QY	121	AspLeuLeuHisAsnValSerGlnAsnIleAlaIleAlaIuThrArgAlaGluAspPro	140
Db	399	GATCTCTTCGATCATCGTGAAGCCAGAAATTTACTCCAAAGACCGGGAGCAGACCATCC	458
QY	141	TrpPheGlu	143
Db	459	TGGTTTGA	467

RESULT 18	
BB628807	643 bp mRNA linear EST-26-OCT-2001
LOCUS	BB628807 RIKEN full-length enriched, 16 days neonate cerebellum Mus musculus; cDNA clone 963003S119 5', mRNA sequence.
DEFINITION	
ACCESSION	BB628807
VERSION	BB628807.1 GI:1646144
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus

Riken integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. *Genome Res.* 10 (11), 1757-1771 (2000)

Komuro, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugihara, Y. and Hayashizaki, Y.

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001)

Kondo, S., Shinagawa, A., Saito, T., Kiyosawa, H., Yamazaki, I., Aizawa, K., Fukuda, S., Hara, A., Itoh, M., Kawai, J., Shibata, K. and Hayashizaki, I.

Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. *Mamm. Genome.* 12, 673-677 (2001)

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES  
source location/Qualifiers

1..643  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="9630035119"  
/clone\_lib="RIKEN full-length enriched, 16 days neonate cerebellum"  
/tissue\_type="cerebellum"  
/dex\_stage="16 days neonate"  
/lab\_host="DH10B"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer 15'  
GAGAGAGAGAGATCCAGAGCTTTTATTTTATTTVN 3', cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot -20.0 and subtraction to Rot -370.0. Second strand cDNA was prepared with the primer adapter of sequence 15' GAGAGAGATTCGAGTTAATTAATATCCCCCCCC 3'. cDNA was cleaved with XhoI and BamHI. Vector: a modified pBluescript KS(+) after bulk excision from Lambda

BASE COUNT 133 a 183 c 195 g 132 t  
ORIGIN

## Alignment Scores:

Pred. No.: 2,07e-51 Length: 643  
Score: 518.50 Matches: 105  
Percent Similarity: 80.13% Conservative: 20  
Best Local Similarity: 67.31% Mismatches: 29  
Query Match: 28.98% Indels: 2  
DB: 10 Gaps: 1

US-09-748-451-2 (1-338) x BB628807 (1-643)

QY 1 MetleuGlnlySProlysSerVallyLeuArgAlaLeuArgSerProArglySPhegly 20  
DB 79 GTGCTCCGCGCAACCAATGCGTCAAGTGGAGCCCTGACATACGCTGCAGATTGCGC 138  
QY 21 ValAlaGlyArgSerGlyGlnGluValLeuArglyGlyCysLeuArgPheGlnLeuPro 40  
DB 139 GTGGCGGCCCGGAGACTGCGAGACTGCTGTAAGGGCTGCTGCTCCAGCTCCCG 198  
QY 41 GlnuArglySerArgLeuGlyLeuArglyGlnuArglyGlnuArglyGlnuArglyPhe 60  
DB 199 ATGCCCGGTTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258  
QY 61 ProSerValProAspAsnAlaGlnuValLeuLeuArglyGlnuArglyGlnuArgly 80  
DB 259 CCGGGCTTCCCAAGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 318  
QY 81 TyrValSerAspIleArgArgPheLeuSerAlaPheHisGlnuArglyGlnuArgly 100  
DB 319 TATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 378  
QY 101 GlnAlaAlaGlnuGlnuLeuGlyCysAspGlnuGlnuAlaProGlnuArglyGlnu 120  
DB 379 CAGGCTCAGCGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 438  
QY 121 AspLeuLeuHisAsnValSerGlnuGlnuAlaAlaGlnu-ThrArgAlaGlnuArgPro 139  
DB 439 GATCTTTTGGATGCTGAGCGCAGAAATATTCTGAGAAACCCGCGGAGAGACCCCA 498  
QY 140 -ProTyrPheGlnuGlnuGlnuArglyGlnuArglyGlnuArglyGlnuArgly 154  
DB 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

DB 499 TCTGCGGTGGAGATCCCAATCTTTTAAACCAAGGGGGG 544

RESULT 19  
BE755721 513 bp mRNA Linear (EST 25-APR-2001)  
LOCUS BE755721  
DEFINITION 209634 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.  
ACCESSION BE755721  
VERSION BE755721.1 GI:10169713  
KEYWORDS EST.  
SOURCE cow.  
ORGANISM Bos taurus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
Bovidae; Bovinae; Bos.  
REFERENCE Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,  
AUTHORS Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,  
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,  
Pettea,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and  
Keefe,J.W.  
Sequence evaluation of four pooled-tissue normalized bovine cDNA  
libraries and construction of a gene index for cattle  
Genome Res. 11 (4), 626-630 (2001)

## TITLE

JOURNALS MEDLINE  
COMMENT

Contact: Smith TPL  
USDA, ARS, US Meat Animal Research Center  
PO Box 166, Clay Center, NE 68933-0166, USA  
Tel: 402 762 4366  
Fax: 402 762 4390

Email: smith@mail.marc.usda.gov  
Single pass sequencing. Bases called and alt\_trimmed with phred  
v0.980904.e. Vector identified by cross-match with the -mnscore 18  
and -minmatch 12 options.  
PCR primers

FORWARD: AGGAACAGCATGACCAT  
BACKWARD: GTTTCAGGTCAGCAGC  
Plate: 59 row: D column: 7  
Seq primer: ATTTAGGACACTATAC.

## FEATURES

source location/Qualifiers

1..513  
/organism="Bos taurus"  
/db\_xref="taxon:9913"  
/clone\_lib="MARC 2BOV"  
/tissue\_type="pooled"  
/lab\_host="DH10B"  
/note="Vector: PCMV SPOR6; Site\_1: NotI; Site\_2: SalI;  
library made from pooled tissue from testis, thymus,  
semiteadonous muscle, longissimus muscle, pancreas,  
adrenal, and endometrium."  
BASE COUNT 90 a 152 c 177 g 94 t  
ORIGIN

## Alignment Scores:

Pred. No.: 3.81e-48 Length: 513  
Score: 490.00 Matches: 96  
Percent Similarity: 83.85% Conservative: 13  
Best Local Similarity: 73.85% Mismatches: 21  
Query Match: 27.39% Indels: 0  
DB: 12 Gaps: 0

US-09-748-451-2 (1-338) x BE755721 (1-513)

QY 1 MetleuGlnlySProlysSerVallyLeuArgAlaLeuArgSerProArglySPhegly 20  
DB 122 GTGCTCCGGAAGCCCAAGAGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 181  
QY 21 ValAlaGlyArgSerGlyGlnGluValLeuArglyGlyCysLeuArgPheGlnLeuPro 40  
DB 182 GTGGCGGGCAGAGACTGCGAGAGAGTCTGCGAAGGGGTGCCAACCTTGCACCTCCC 241  
QY 41 GlnuArglySerArgLeuGlyLeuArglyGlnuArglyGlnuArglyGlnuArglyPhe 60  
DB 242 ATCCCTGGTCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 301



THIS PAGE BLANK